

Inv: SOBek et al.

CRFE

U.S. DEPARTMENT OF COMMERCE  
Patent and Trademark Office

R Date: 3/9/01

# SEARCH REQUEST FORM

113445

Requestor's Name: JANE ZARA

Serial Number: 09/803,165

Date: 2-6-04

Phone: 2-0765

Art Unit: 1635

2018

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

11/81

Please Search Seq ID No: 34

No size limits

Thanks

## STAFF USE ONLY

Date completed: 2/10  
Searcher: \_\_\_\_\_  
Terminal time: \_\_\_\_\_  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: \_\_\_\_\_  
Number of Searches: \_\_\_\_\_  
Number of Databases: \_\_\_\_\_

### Search Site

\_\_\_\_ STIC  
\_\_\_\_ CM-1  
\_\_\_\_ Pre-S

### Type of Search

\_\_\_\_ N.A. Sequence  
\_\_\_\_ A.A. Sequence  
\_\_\_\_ Structure  
\_\_\_\_ Bibliographic

### Vendors

\_\_\_\_ IG  
\_\_\_\_ STN  
\_\_\_\_ Dialog  
\_\_\_\_ APS  
\_\_\_\_ Geninfo  
\_\_\_\_ SDC  
\_\_\_\_ DARC/Questel  
\_\_\_\_ Other

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2004, 11:48:25 ; Search time 28 Seconds  
(without alignments)  
1169.593 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026

Sequence: 1 MIFDIDYITKGRPIIRPK.....KEDLKYSKQVGLDNLWK 774

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCUTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	3225.5	80.1	774	3	US-08-902-632-2
2	3222.5	80.0	774	3	US-09-073-354-1
3	3222.5	80.0	774	3	US-08-656-005A-1
4	3222.5	80.0	774	3	US-09-073-259-1
5	3222.5	80.0	774	3	US-09-363-095-1
6	3222.5	80.0	774	3	US-09-418-027-1
7	3209.5	79.7	778	2	US-08-906-925-4
8	3151	78.3	775	1	US-07-966-278-1
9	3151	78.3	775	1	US-08-424-921-1
10	3151	78.3	775	2	US-08-556-355A-1
11	3151	78.3	775	2	US-07-803-627A-1
12	3151	78.3	775	4	US-09-244-889A-1
13	3151	78.3	776	2	US-08-688-649-37
14	3141.5	78.0	773	4	US-09-585-858-35
15	3139	78.0	779	1	US-08-375-134-12
16	3139	78.0	779	5	PCT-US95-15263-12
17	2974	73.9	727	4	US-09-585-858-32
18	2502.5	62.2	1022	1	US-08-271-364A-8
19	2502.5	62.2	1022	2	US-08-222-715B-27
20	2069	51.4	1019	1	US-08-271-364A-7
21	2069	51.4	1019	2	US-08-222-715B-26
22	1595	39.6	396	1	US-08-229-284A-2
23	1403	34.8	788	4	US-08-907-166-6
24	1403	34.8	788	4	US-09-391-340-6
25	1194	29.7	803	2	US-08-907-166-4
26	1194	29.7	803	4	US-09-391-340-4
27	1191	29.6	803	1	US-08-062-368-4

28	1180	29.3	803	1	US-08-062-368-2
29	656	16.3	734	4	US-09-585-858-8
30	611	15.2	738	4	US-09-585-858-7
31	611	15.2	1107	1	US-08-366-577-2
32	611	15.2	1107	5	PCT-US96-00005-2
33	602	15.0	735	4	US-09-585-858-10
34	602	15.0	738	4	US-09-585-858-6
35	599.5	14.9	734	4	US-09-585-858-9
36	591	14.7	877	2	US-08-907-166-8
37	591	14.7	877	4	US-09-391-340-8
38	588.5	14.6	762	2	US-08-907-166-10
39	588.5	14.6	762	4	US-09-391-340-10
40	585.5	14.5	702	4	US-09-585-858-33
41	585.5	14.5	1462	3	US-07-792-600-31
42	585.5	14.5	1462	3	US-09-157-021-31
43	585.5	14.5	1462	3	US-09-156-842-31
44	582	14.5	757	4	US-09-585-858-23
45	568.5	14.1	757	4	US-09-585-858-24

#### ALIGNMENTS

RESULT 1  
US-08-902-632-2  
; Sequence 2, Application US/08902632  
; Patent No. 6008025  
; GENERAL INFORMATION:  
; APPLICANT: KOMATSUBARA, Hideyuki  
; APPLICANT: KITABAYASHI, Masao  
; APPLICANT: KAMIMURA, Hideki  
; APPLICANT: KAWAKAMI, Bunsei  
; APPLICANT: KAWAMURA, Yoshihisa  
; APPLICANT: TAKAGI, Masahiro  
; APPLICANT: IMANAKA, Tadayuki  
; TITLE OF INVENTION: Modified Thermostable DNA Polymerase,  
; TITLE OF INVENTION: and DNA Polymerase Composition for Nucleic Acid  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1 Broadway  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10004

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.50 inch floppy disk, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/902,632  
FILING DATE: Concurrent Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 198911/96  
FILING DATE: 29-JUL-1996  
APPLICATION NUMBER: JP 200446/96  
APPLICATION NUMBER: 30-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: GREASON, Edward W.  
REGISTRATION NUMBER: 18,918  
REFERENCE/DOCKET NUMBER: 2418/7  
TELEPHONE: 212-425-7200  
TELEFAX: 212-425-5288  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 774 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-902-632-2

Query Match 80.1%; Score 3225.5; DB 3; Length 774;  
Best Local Similarity 77.5%; Pred. No. 1.2e-242;  
Matches 599; Conservative 82; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIFDTDYITKDGKPIIRIFKKGNEGFKIELDPHPQPIYALLKDDSAIDBIKAIKGERHG 60  
DB 1 MILDTDYITDGKPIIRIFKKGNEGFKIEYDRTFEPFYALLKDDSAIEVKKITAERHG 60

QY 61 KIVRVVDVAVKVKKFLGRDVEVWKLFEHPQDVPAIRGKIREHPAVIDIYEDIDIPAKRY 120  
DB 61 TVVTVKRVEKVKQKFLGRDVEVWKLFEHPQDVPAIRGKIREHPAVIDIYEDIDIPAKRY 120

QY 121 LIDKGLIPWEGDEELKMAFDIETFYHGEDEFGKEIIMISYADESEARVITWKNIDLPY 180  
DB 121 LIDKGLVPMWEGDEELKMAFDIETFYHGEDEFGKEIIMISYADESEARVITWKNVDLPY 180

QY 181 VDVVSNEREMIKRFLVQIVREKDPDVLITYNGDNFDLPYLKRAEKLGVTLGLGRDKHPE 240  
DB 181 VDVVSNEREMIKRFLVQIVREKDPDVLITYNGDNFDLPYLKRAEKLGVTLGLGRDKHPE 240

QY 241 PKIHRMGDSFAVEIKRTHFDLPVVRRTINLPTTYLGAUVEAVLKGKSKLGAEEIAAI 300  
DB 239 PKIQMGDRFAVEIKRTHFDLPVVRRTINLPTTYLGAUVEAVLKGKSKLGAEEIATPA 298

QY 301 WETESMKLAQYSMEDARATYELGKEFPFMEALAKLIGOSVWDVSRSSSTGNLWVYLL 360  
DB 299 WETGENLERVARYSMEDAKVYELGKEFPFMEALAKLIGOSLWVSRSSSTGNLWVFL 358

QY 361 RVAYERNELAPNKPDEEYRRLRTTYLGGVYKPEPGLWENIYLDPRCLYPSIIITHN 420  
DB 359 RKAYERNELAPNKPDEEYRRLRTTYLGGVYKPEPGLWENIYLDPRCLYPSIIITHN 417

QY 421 VSPDTLREGCKNVDVAPIGVYCKDFPQFIISILGELITMQEIKKWKATIDPLEKK 480  
DB 418 VSPDTLREGCKNVDVAPIGVYCKDFPQFIISILGELITMQEIKKWKATIDPIERK 477

QY 481 MLDYQRAVKLHANSYGYGYPKARWYKCAESVTWANGRYHMTIETKEEKGKVL 540  
DB 478 LLDYQRAVKLHANSYGYGYPKARWYKCAESVTWANGRYHMTIETKEEKGKVKI 537

QY 541 YADTQGFVATIPGKPEITKKAKKFLKINSKLPGLLEVEYEGFYLRFVAKKRYAVI 600  
DB 538 YSDTQGFVATIPGKPEITKKAKKFLKINSKLPGLLEVEYEGFYLRFVAKKRYAVI 597

QY 601 DEGRITTRGLEVRDWESEIAKETQAKVLEBAILKEDSVEKAVEIKVDVVEIAKYQVPL 660  
DB 598 DEGRITTRGLEVRDWESEIAKETQAKVLEBAILKEDSVEKAVEIKVDVVEIAKYQVPP 657

QY 661 EKLVIHQITKDLSEYKAGPHVAIAKRLAAKIGKVRPGTITSYIVLRGSGKISDRVILL 720  
DB 658 EKLVIHQITRDLKDYKAGPHVAIAKRLAAKIGKVRPGTITSYIVLRGSGRIGDRAIPF 717

QY 721 SEYDPKHKYDPDYIENQVLPVLRILEAFGRKEDLUKQSKOVGLDAMUK 773  
DB 718 DBFDPFKHKYDAEYIENQVLPVLRILEAFGRKEDLUKQSKOVGLDAMUK 770

RESULT 2  
US-09-073-354-1

; Sequence 1, Application US/09073354  
; Patent No. 6033859  
; GENERAL INFORMATION:  
; APPLICANT: KITABAYASHI, Masao  
; APPLICANT: ARAKAWA, Taku  
; APPLICANT: INOUE, Hiroaki  
; APPLICANT: KAWAKAMI, Bunsei  
; APPLICANT: KAWAMURA, Yoshihisa  
; APPLICANT: IMANAKA, Tadayuki  
; APPLICANT: TAKAGI, Masahiro  
; APPLICANT: MORIKAWA, Masaaki  
; TITLE OF INVENTION: A Thermostable DNA Polymerase and Kits for

TITLE OF INVENTION: Amplifying Nucleic Acids  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073.354  
FILING DATE: Concurrent Herewith  
CLASSIFICATION: 0506  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/656,005  
FILING DATE: 24 MAY 1996  
APPLICATION NUMBER: JP 134096/95  
FILING DATE: 31 MAY 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 2418/9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 774 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-073-354-1

Query Match 80.0%; Score 3222.5; DB 3; Length 774;  
Best Local Similarity 77.4%; Pred. No. 2e-242;  
Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIFDTDYITKDGKPIIRIFKKGNEGFKIELDPHPQPIYALLKDDSAIDBIKAIKGERHG 60  
DB 1 MILDTDYITDGKPIIRIFKKGNEGFKIEYDRTFEPFYALLKDDSAIEVKKITAERHG 60

QY 61 KIVRVVDVAVKVKKFLGRDVEVWKLFEHPQDVPAIRGKIREHPAVIDIYEDIDIPAKRY 120  
DB 61 TVVTVKRVEKVKQKFLGRDVEVWKLFEHPQDVPAIRGKIREHPAVIDIYEDIDIPAKRY 120

QY 121 LIDKGLIPWEGDEELKMAFDIETFYHGEDEFGKEIIMISYADESEARVITWKNIDLPY 180  
DB 121 LIDKGLVPMWEGDEELKMAFDIETFYHGEDEFGKEIIMISYADESEARVITWKNVDLPY 180

QY 181 VDVVSNEREMIKRFLVQIVREKDPDVLITYNGDNFDLPYLKRAEKLGVTLGLGRDKHPE 240  
DB 181 VDVVSNEREMIKRFLVQIVREKDPDVLITYNGDNFDLPYLKRAEKLGVTLGLGRDKHPE 240

QY 241 PKIHRMGDSFAVEIKRTHFDLPVVRRTINLPTTYLGAUVEAVLKGKSKLGAEEIAAI 300  
DB 239 PKIQMGDRFAVEIKRTHFDLPVVRRTINLPTTYLGAUVEAVLKGKSKLGAEEIATPA 298

QY 301 WETESMKLAQYSMEDARATYELGKEFPFMEALAKLIGOSVWDVSRSSSTGNLWVYLL 360  
DB 299 WETGENLERVARYSMEDAKVYELGKEFPFMEALAKLIGOSLWVSRSSSTGNLWVFL 358

QY 361 RVAYERNELAPNKPDEEYRRLRTTYLGGVYKPEPGLWENIYLDPRCLYPSIIITHN 420  
DB 359 RKAYERNELAPNKPDEEYRRLRTTYLGGVYKPEPGLWENIYLDPRCLYPSIIITHN 417

QY 421 VSPDTLREGCKNVDVAPIGVYCKDFPQFIISILGELITMQEIKKWKATIDPLEKK 480  
DB 418 VSPDTLREGCKNVDVAPIGVYCKDFPQFIISILGELITMQEIKKWKATIDPIERK 477

Db 418 VSPDTLNREGCKEYDVAPQVGHGFCKDPFGFIPSLGLDLLEERQKIKKKMKATIDPIERK 477  
Qy 481 MLDYRQRAVKLHANSYGYGVPKARWYKSCASVTAWGRHYLWMTIKETEERKFGKVL 540  
Db 478 LLDYRQRAIKILANSYGYGVARARVCKCAESVTAWGREYITMTIKETEERKFGKVI 537  
Qy 541 YADTDGFTATIPGADAEVTKKAMEFLNYINAKLPGALEVEYEGFYKRGFPVTKKYAVI 600  
Db 538 YSDTDGFTATIPGADAEVTKKAMEFLNYINAKLPGALEVEYEGFYKRGFPVTKKYAVI 597  
Qy 601 DEEGRIITRGLVVRDSEIAKETQAKVLEAILKEDSVKAVELKVDVEZIEIAKYOVPL 660  
Db 598 DEEGRIITRGLVVRDSEIAKETQAKVLEAILKEDSVKAVELKVDVEZIEIAKYOVPL 657  
Qy 661 EKLVIHEQITKOLSEYKAIGPHVAIAKRLAAKGIVKRPOTIISYIVLRGSKISDRVILL 720  
Db 658 EKLVIHEQITRDLKDYKATGPHVAVAKRLAAARGVKIRPGTVISYIVLRGSKIGDRAIFF 717  
Qy 721 SEYDPKHKYDPPDYIENQVLPVAILPAFGRKEDLKYQSSKQVGLDANLK 773  
Db 718 DEFDPKHKYDAEYIENQVLPVAILPAFGRKEDLKYQSSKQVGLDANLK 770

## RESULT 3

US-08-656-005A-1  
; Sequence 1, Application US/08656005A  
; Patent No. 6054301  
; GENERAL INFORMATION:  
; APPLICANT: KITABAYASHI, Masao  
; APPLICANT: ARAKAWA, Taku  
; APPLICANT: INOUE, Hiroaki  
; APPLICANT: KAWAKAMI, Bunsei  
; APPLICANT: KAWAMURA, Yoshihisa  
; APPLICANT: IWANAKA, Tadayuki  
; APPLICANT: TAKAGI, Masahiro  
; APPLICANT: MORIKAWA, Masaaki  
; TITLE OF INVENTION: A Method of Amplifying Nucleic  
; TITLE OF INVENTION: Acid and A Reagent Therefor  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/656,005A  
; FILING DATE: 24 MAY 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 134096/95  
; FILING DATE: 31 MAY 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 2418/3  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 774 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-656-005A-1

Query Match 80.0%; Score 3222.5; DB 3; Length 774;  
Best Local Similarity 77.4%; Pred No 2a-242;  
Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

Qy 1 MIPDITITKQKPIIRIPFKKENGSKFELDHPHOFYIYALLKDDSAIDIEIKAIKERHG 60  
Db 1 MILDITITEDGKPVIRIPFKKENGSKFIEYDTPFPFYALLKDDSAIEVKKITAEHRG 60  
Qy 61 KIVRVVDAVKKKFLGRDVEVWKLIFHPQDPVAPLGRKIRREHPAVIDIYVDIPAKY 120  
Db 61 TVTVTRVREKVKKFLGRFPVEWKLFTHPQDPVAPLGRKIRREHGAVIDIYVDIPAKY 120  
Qy 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKGEIIMISYADSEEAARVITWKNIDLPY 180  
Db 121 LIDKGLVPMGEDEELKMAFDIETFYHGEDEFGKGEIIMISYADSEEAARVITWKNIDLPY 180  
Qy 181 VDVSNEREMIKRFVQIVREKDPDVLITVNGDNFDPYLIKRAEKLGVTLGLGRDKEHE 240  
Db 181 VDVSNEREMIKRFVQIVREKDPDVLITVNGDNFDPYLIKRAEKLGVTLGLGRDKEHE 240  
Qy 241 PKIHRMGDSFAVEIKGRIHFDLPFVVRRTINLPTTLEAVYEAVLGKTCKSLGABEIAAI 300  
Db 239 PKIORMGDRFAVEIKGRIHFDLPFVVRRTINLPTTLEAVYEAVLGKTCKSLGABEIAAI 298  
Qy 301 WETERSMKKLAQYSEMEDAATYELGKEFPFMEALAKLIGOSVWVSSTGNLWVWILL 360  
Db 299 WETGENLVARVSEMEDAKVTYELGKEFLPMEASQSLRLLIGOSLWVVSSTGNLWVWILL 358  
Qy 361 RVAYERNELAPNKPDEEYRRRLRTYLGYYKVEPERGLWENITTYLDRCFLYPSIIVTHN 420  
Db 359 RKAYERNELAPNKPDEKELARR-RQSYEGGYKVEPERGLWENITTYLDRCFLYPSIIVTHN 417  
Qy 421 VSPDTLEREGCKYDVAPIGVYKFCDFPGFIPSLIGELITWROETKMKKATIDIEKK 480  
Db 418 VSPDTLNREGCKEYDVAPQVGHGFCKDPFGFIPSLGLDLLEERQKIKKKMKATIDPIERK 477  
Qy 481 MLDYRQRAVKLHANSYGYGVPKARWYKSCASVTAWGRHYLWMTIKETEERKFGKVL 540  
Db 478 LLDYRQRAIKILANSYGYGVARARVCKCAESVTAWGREYITMTIKETEERKFGKVI 537  
Qy 541 YADTDGFTATIPGADAEVTKKAMEFLNYINAKLPGALEVEYEGFYKRGFPVTKKYAVI 600  
Db 538 YSDTDGFTATIPGADAEVTKKAMEFLNYINAKLPGALEVEYEGFYKRGFPVTKKYAVI 597  
Qy 601 DEEGRIITRGLVVRDSEIAKETQAKVLEAILKEDSVKAVELKVDVEZIEIAKYOVPL 660  
Db 598 DEEGRIITRGLVVRDSEIAKETQAKVLEAILKEDSVKAVELKVDVEZIEIAKYOVPL 657  
Qy 661 EKLVIHEQITKOLSEYKAIGPHVAIAKRLAAKGIVKRPOTIISYIVLRGSKISDRVILL 720  
Db 658 EKLVIHEQITRDLKDYKATGPHVAVAKRLAAARGVKIRPGTVISYIVLRGSKIGDRAIFF 717  
Qy 721 SEYDPKHKYDPPDYIENQVLPVAILPAFGRKEDLKYQSSKQVGLDANLK 773  
Db 718 DEFDPKHKYDAEYIENQVLPVAILPAFGRKEDLKYQSSKQVGLDANLK 770

## RESULT 4

US-09-073-259-1  
; Sequence 1, Application US/09073259  
; Patent No. 6143536  
; GENERAL INFORMATION:  
; APPLICANT: IMANAKA, Tadayuki  
; APPLICANT: TAKAGI, Masahiro  
; APPLICANT: MORIKAWA, Masaaki  
; TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.

```

/ COUNTRY: USA
/ ZIP: 20036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: WordPerfect 6.1 Windows
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/073,259
/ FILING DATE: Concurrent Herewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/656,005
/ FILING DATE: 24 MAY 1996
/ APPLICATION NUMBER: JP 134096/95
/ FILING DATE: 31 MAY 1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Toffenetti, Judith L.
/ REGISTRATION NUMBER: 39,048
/ REFERENCE/DOCKET NUMBER: 2418/10
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-429-1776
/ TELEFAX: 202-429-0796
/ INFORMATION FOR SEQ ID NO. 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 774 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-073-259-1

Query Match 80.0%; Score 3222.5; DB 3; Length 774;
Best Local Similarity 77.4%; Pred. No. 2e-242;
Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIFDTYITKDGKPIIRIFKKGEGFKIELDPHFQPIYIALLKDDSAIDRIKAKGERHG 60
DB 1 MILDYITEDGKPIIRIFKKGEGFKIYDRTPEFYALLKDDSAIEVKKIATERRHG 60
QY 61 KIVRVVDAVKVKKLGRODVEVWKLIFEPDQVPALRGKIREHPAVIDIYEDIPFAKRY 120
DB 61 TVTVTVKRVKVKKLGRODVEVWKLIFEPDQVPALRGKIREHPAVIDIYEDIPFAKRY 120
QY 121 LIDKGLIPMEGDEELKMAFDIETPYHEGDFGKEGIMISYADEEGARVITWKNIDLPY 180
DB 121 LIDKGLVPMGDEELKMAFDIOTLYHEGEFAEGPILMISYADEEGARVITWKNVDLPY 180
QY 181 VDVSNEREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKKEAKLGVTLLGRDKEHPE 240
DB 181 VDVSNEREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKKEAKLGVTLLGRDKEHPE 240
QY 241 PKIHRMGDSFAVEIKGRIHFDLPFPVVRTINLPTTYLLEAVYEAVALGKTSKLGAEETAAI 300
DB 239 PKIQRMGDRFAVEIKGRIHFDLPFPVVRTINLPTTYLLEAVYEAVALGKTSKLGAEETAAI 300
QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEALKIGOSVNDVSRSSSTGNLVEWYLL 360
DB 299 WETGENLERVARYSMEDAKVTYELGKEFFPMEAEALKIGOSVNDVSRSSSTGNLVEWYLL 360
QY 361 RVAYERNELAPNKPDDEEYRRLRTTYLGGYVKEPERGLMENITYLDFRCLYPSIIVTHN 420
DB 359 RKAYERNELAPNKPDDEEYRRLRTTYLGGYVKEPERGLMENITYLDFRCLYPSIIVTHN 420
QY 421 VSPDTLRGCKNDYAPVGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 480
DB 418 VSPDTLRGCKNDYAPVGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 480
QY 481 MLYRQRVAKVLANSGYGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 540
DB 478 MLYRQRVAKVLANSGYGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 540
QY 541 YADTDGFFATIPGADAEIVTKKAMEFLNYNAKLPALBLEYEGFYKRGFFVTKKXYAVI 597
DB 538 YADTDGFFATIPGADAEIVTKKAMEFLNYNAKLPALBLEYEGFYKRGFFVTKKXYAVI 597
QY 601 DEEGRIITRGLVVRDWSIAKETQAKVLEAILKEDSVKEKAVEIVKDVVBEIAKQVPL 660
DB 598 DEEGRIITRGLVVRDWSIAKETQAKVLEAILKEDSVKEKAVEIVKDVVBEIAKQVPL 660
QY 661 EKLVIHQITKDLSEYKAIGPHVAIAKRLAAKGIKVRPGTIISYIVLRSGSKISDRVILL 720
DB 658 EKLVIHQITKDLSEYKAIGPHVAIAKRLAAKGIKVRPGTIISYIVLRSGSKISDRVILL 720
QY 721 SEYDPKXHKYDPPDYIENQVLPVLRILEAFGYKREDLKYQSSKQVGLDAMLK 773
DB 718 SEYDPKXHKYDPPDYIENQVLPVLRILEAFGYKREDLKYQSSKQVGLDAMLK 773

RESULT 5
US-09-363-095-1
; Sequence 1, Application US/09363095
; Patent No. 6187573
; GENERAL INFORMATION:
; APPLICANT: IMANAKA, Tadayuki
; APPLICANT: TAKAGI, Masahiro
; APPLICANT: MORIKAWA, Masaaki
; TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase
; FILE REFERENCE: 2418/11
; CURRENT APPLICATION NUMBER: US/09/363,095
; CURRENT FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/073,259
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: US 08/656,005
; EARLIER FILING DATE: 1996-05-24
; EARLIER APPLICATION NUMBER: JP 134096/95
; EARLIER FILING DATE: 1995-05-31
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: WordPerfect 6.1 Windows
; SEQ ID NO 1
; TYPE: PRT
; LENGTH: 774
; ORGANISM: Hyperthermophilic archaeon
; US-09-363-095-1

Query Match 80.0%; Score 3222.5; DB 3; Length 774;
Best Local Similarity 77.4%; Pred. No. 2e-242;
Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIFDTYITKDGKPIIRIFKKGEGFKIELDPHFQPIYIALLKDDSAIDRIKAKGERHG 60
DB 1 MILDYITEDGKPIIRIFKKGEGFKIYDRTPEFYALLKDDSAIEVKKIATERRHG 60
QY 61 KIVRVVDAVKVKKLGRODVEVWKLIFEPDQVPALRGKIREHPAVIDIYEDIPFAKRY 120
DB 61 TVTVTVKRVKVKKLGRODVEVWKLIFEPDQVPALRGKIREHPAVIDIYEDIPFAKRY 120
QY 121 LIDKGLIPMEGDEELKMAFDIETPYHEGDFGKEGIMISYADEEGARVITWKNIDLPY 180
DB 121 LIDKGLVPMGDEELKMAFDIOTLYHEGEFAEGPILMISYADEEGARVITWKNVDLPY 180
QY 181 VDVSNEREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKKEAKLGVTLLGRDKEHPE 240
DB 181 VDVSNEREMIKRFVQIVREKDPDVLITVNGDNFDLPYLKKEAKLGVTLLGRDKEHPE 240
QY 241 PKIHRMGDSFAVEIKGRIHFDLPFPVVRTINLPTTYLLEAVYEAVALGKTSKLGAEETAAI 300
DB 239 PKIQRMGDRFAVEIKGRIHFDLPFPVVRTINLPTTYLLEAVYEAVALGKTSKLGAEETAAI 300
QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEALKIGOSVNDVSRSSSTGNLVEWYLL 360
DB 299 WETGENLERVARYSMEDAKVTYELGKEFFPMEAEALKIGOSVNDVSRSSSTGNLVEWYLL 360
QY 361 RVAYERNELAPNKPDDEEYRRLRTTYLGGYVKEPERGLMENITYLDFRCLYPSIIVTHN 420
DB 359 RKAYERNELAPNKPDDEEYRRLRTTYLGGYVKEPERGLMENITYLDFRCLYPSIIVTHN 420
QY 421 VSPDTLRGCKNDYAPVGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 480
DB 418 VSPDTLRGCKNDYAPVGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 480
QY 481 MLYRQRVAKVLANSGYGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 540
DB 478 MLYRQRVAKVLANSGYGVGKFCDFPFGFTPSLIGLITWRQIKKKMATIDPLEKK 540
QY 541 YADTDGFFATIPGADAEIVTKKAMEFLNYNAKLPALBLEYEGFYKRGFFVTKKXYAVI 597
DB 538 YADTDGFFATIPGADAEIVTKKAMEFLNYNAKLPALBLEYEGFYKRGFFVTKKXYAVI 597

```

```

Qy 421 VSPDTLREGCNVDVAPIVGYKFCDFPGFIPISILGELITMQEIKKKOMKATIDPIEK 480
Db 418 VSPDTLREGCNVDVAPIVGYKFCDFPGFIPISILGELITMQEIKKKOMKATIDPIEK 477
Qy 481 MLDYRQRAVKLHANSYGYGMYGPKARWYKCECAESVTAWGRHYIEMTIKIEIKFGPKVL 540
Db 478 LLDYRQRAIKLANSYGYGYARARWYKCECAESVTAWGREYITWTIKIEIKFGPKVI 537
Qy 541 YADTDGYATIPGKPTTIKKAKFELKYINSKLPGLLEVEGFLRGFFVAKKRYAVI 600
Db 538 YSDTDGFATIPGADAEIVKKAMEFLNYINAKLPGLALEVEGFLRGFFVTKKYAVI 597
Qy 601 DEGRITTRGLEIVRRDWSIAKETQAKVLEAILKEDSVKAVEIVKQVVBETAKYQVPL 660
Db 598 DEGRITTRGLEIVRRDWSIAKETQAKVLEAILKEDSVKAVEIVKQVVBETAKYQVPL 657
Qy 661 EKLVIHQITKDLSEYKAIGHVAIAKRLAAKGIKVRPGTIISYILRGSGKISDRVILL 720
Db 658 EKLVIHQITRDLKDYKATGPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIGDRAIFP 717
Qy 721 SEYDPKXHKYDPPDYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWLK 773
Db 718 DEFDPKHKYDAEYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWLK 770

```

## RESULT 6

```

; US-09-418-027-1
; Sequence 1, Application US/09418027
; Patent No. 6225065
; GENERAL INFORMATION:

```

```

; APPLICANT: KITABAYASHI, Masao
; APPLICANT: ARAKAWA, Taku
; APPLICANT: INOUE, Hiroaki
; APPLICANT: KAWAKAMI, Bunsei
; APPLICANT: KAWAMURA, Yoshihisa
; APPLICANT: IMANAKA, Tadayuki
; APPLICANT: TAKAGI, Masahiro
; APPLICANT: MORIKAWA, Masaaki

```

```

; TITLE OF INVENTION: A Thermostable DNA Polymerase and Kits for
; Amplifying Nucleic Acids

```

```

; NUMBER OF SEQUENCES: 16

```

```

; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Kenyon & Kenyon

```

```

; STREET: 1025 Connecticut Avenue, N.W., Suite 600

```

```

; CITY: Washington

```

```

; STATE: D.C.

```

```

; COUNTRY: USA

```

```

; ZIP: 20036

```

```

; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: 3 5" Floppy disk

```

```

; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: WordPerfect 6.1 Windows

```

```

; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/418,027

```

```

; FILING DATE:

```

```

; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 09/073,354

```

```

; FILING DATE:

```

```

; APPLICATION NUMBER: JP 134096/95

```

```

; FILING DATE: 31 MAY 1995

```

```

; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Toffenetti, Judith L.

```

```

; REGISTRATION NUMBER: 39,046

```

```

; REFERENCE/DOCKET NUMBER: 2418/9

```

```

; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 202-429-1776

```

```

; TELEFAX: 202-429-0796

```

```

; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 774 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-418-027-1

```

```

Query Match 80.0%; Score 3222.5; DB 3; Length 774;

```

```

Best Local Similarity 77.4%; Pred. No. 2e-242;
Matches 598; Conservative 83; Mismatches 99; Indels 3; Gaps 2;

```

```

Qy 1 MIFDNDYITKDGKPIIRIFKXENGEFKIELDPHFQPIYVALLKDDSAIDEIKAIKGEHG 60
Db 1 MLDTDYITDYGKPIRIFKXENGEFKIELDPHFQPIYVALLKDDSAIDEIKAIKGEHG 60
Qy 61 KIVRVDAVKKPLGRDVEVWKLIFHPQDVPAALRGKIREHPAVDIYDIYDIPFAKRY 120
Db 61 TVTVTKRVEKQKPLGRDVEVWKLIFHPQDVPAALRGKIREHPAVDIYDIYDIPFAKRY 120
Qy 121 LIDKGLIPMEGDEELKMAFDIETFYHEGDFGKEIIMISYADEEARVITWKNIDLPY 180
Db 121 LIDKGLIPMEGDEELKMAFDIOTLYHEGEFAGPILMISYADEEGARVITWKNIDLPY 180
Qy 181 VDVVSNREMIKRFVQIVREKDPDLITVINGDNFDLPYLKRAEKLGVTLGLGRDKEHPE 240
Db 181 VDVVSTEREMIKRFLRVVKEKDPDLITVINGDNFDLPYLKRAEKLGVTLGLGRDGS--E 238
Qy 241 PKIHRMGDSFAVEIKGRITHFDLPVVRTINLPYTLVAVYVAVLTKSKLGAEEIAAI 300
Db 239 PKIQMGDRFAVEIKGRITHFDLPVVRTINLPYTLVAVYVAVYVAVLTKSKLGAEEITPA 298
Qy 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEAKLIGQSVWDVSRSSSTGNLVEWYLL 360
Db 299 WETGENLVARVYSMEDAKVYELGKEFLPMEAQSLRIGQSLWDVSRSSSTGNLVEWFL 358
Qy 361 RVAVERNELAPNKDEBEYERELRTTVYLGKVEPERGLMENITVLPCLYPSIIVTHN 420
Db 359 RRAVERNELAPNKDEBEYERELRTTVYLGKVEPERGLMENITVLPCLYPSIIVTHN 417
Qy 421 VSPDTLREGCNVDVAPIVGYKFCDFPGFIPISILGELITMQEIKKKOMKATIDPIEK 480
Db 418 VSPDTLREGCNVDVAPIVGYKFCDFPGFIPISILGELITMQEIKKKOMKATIDPIEK 477
Qy 481 MLDYRQRAVKLHANSYGYGMYGPKARWYKCECAESVTAWGRHYIEMTIKIEIKFGPKVL 540
Db 478 LLDYRQRAIKLANSYGYGYARARWYKCECAESVTAWGREYITWTIKIEIKFGPKVI 537
Qy 541 YADTDGYATIPGKPTTIKKAKFELKYINSKLPGLLEVEGFLRGFFVAKKRYAVI 600
Db 538 YSDTDGFATIPGADAEIVKKAMEFLNYINAKLPGLALEVEGFLRGFFVTKKYAVI 597
Qy 601 DEGRITTRGLEIVRRDWSIAKETQAKVLEAILKEDSVKAVEIVKQVVBETAKYQVPL 660
Db 598 DEGRITTRGLEIVRRDWSIAKETQAKVLEAILKEDSVKAVEIVKQVVBETAKYQVPL 657
Qy 661 EKLVIHQITKDLSEYKAIGHVAIAKRLAAKGIKVRPGTIISYILRGSGKISDRVILL 720
Db 658 EKLVIHQITRDLKDYKATGPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIGDRAIFP 717
Qy 721 SEYDPKXHKYDPPDYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWLK 773
Db 718 DEFDPKHKYDAEYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWLK 770

```

## RESULT 7

```

; US-08-906-925-4
; Sequence 4, Application US/08906925
; Patent No. 5882904
; GENERAL INFORMATION:

```

```

; APPLICANT: Riedl, William

```

```

; APPLICANT: Fly, Susan

```

```

; APPLICANT: Kaboord, Barbara F.

```

```

; APPLICANT: Nye, Steven H.

```

```

; APPLICANT: Ting, Eve

```

/ TITLE OF INVENTION: THERMOCOCCUS BAROSII DNA POLYMERASE MUTANTS  
 / NUMBER OF SEQUENCES: 9  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Quarles & Brady  
 / STREET: 411 East Wisconsin Avenue  
 / CITY: Milwaukee  
 / STATE: Wisconsin  
 / COUNTRY: U.S.A.  
 / ZIP: 53202-4497  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patent in Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/906,925  
 / FILING DATE:  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Baker, Jean C.  
 / REGISTRATION NUMBER: 35,433  
 / REFERENCE/DOCKET NUMBER: 740211.90628  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (414) 277-5709  
 / TELEFAX: (414) 271-3552  
 / INFORMATION FOR SEQ ID NO: 4:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 778 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-08-906-925-4

Query Match 79.7%; Score 3209.5; DB 2; Length 778;  
 Best Local Similarity 77.2%; Pred. No. 2.1e-241;  
 Matches 596; Conservative 93; Mismatches 90; Indels 3; Gaps 2;

QY 1 MIFDYYITKDGKPIIRIFKENGEGFKIELDPHFQPIYIALLKDDSAIDEIKAKGERHG 60  
 DB 1 MILVDYITTEDGKPIRVFKKDGFKIEYDFEPFIYALLRDSASIEIEKITAERHG 60  
 QY 61 KIVRVVDAVKKKKGLGRVVEWKLIFHPQVDPALRGKIREHPAVIDIYDIPFAKRY 120  
 DB 61 KVVVKRAEKVKKKKGLGRSVVWVLYFHPQVPAIRDKIRKHPAVIDIYDIPFAKRY 120  
 QY 121 LIDKGLIPMEGDEELKMAFDIETPFHGEDEGKGEIIMISYADDEEARVITWKNIDLPY 180  
 DB 121 LIDKGLVPNEGDEELKMSFDIETLYHEGEERGTGPILMISYADSEARVITWKNIDLPY 180  
 QY 181 VDVSNEREMIKRFVQIVREKQPDVLTITNGDNFDPYLIIRAEKGLVTLILGRDKEHPE 240  
 DB 181 VDVVSTEKEMIKRFLKVVKEKQPDVLTITNGDNFDPYLIIRAEKGLVSTLGRDGS--E 238  
 QY 241 PKIHRMGDSFAVEIKGRIHFDLPVVRRTINLPTTLEAVYBAVLGKTKSLGAEIAAI 300  
 DB 239 PKIQMGDRFAVEIKGRIHFDLPVVRRTINLPTTLEAVYBAVLGKTKSLGAEIAAI 298  
 QY 301 WETESMKLAQYSDMADATYELGKEFPFPPMAELKIGQSVWDVSRSSSTGNLVWYLL 360  
 DB 299 WETGELRVARYSDMADATYELGKEFPFPPMAELKIGQSVWDVSRSSSTGNLVWYLL 358  
 QY 361 RVAYERNELAPNKPDEEYRRLRTYLLGGYKKEPERGLWENITVLDRCYPSIIVTHN 420  
 DB 359 RKAYERNELAPNKPDEEYRRLRTYLLGGYKKEPERGLWENITVLDRCYPSIIVTHN 417  
 QY 421 VSPDTLREGCKNYDVAPITGVYKFKDPPGFIPIISILGELITWROBKIKKMKATIDPIBK 480  
 DB 418 VSPDTLNRGCKSYDVAPQVGHKFKDPPGFIPIISILGELITWROBKIKKMKATIDPIBK 477  
 QY 481 MLDYRORAKLHANSYGVYKPKARWYSKECAESVTWAGRYIEMTIKEEKEGFKVL 540  
 DB 478 LLDYRORAKLHANSYGVYKPKARWYSKECAESVTWAGRYIEMTIKEEKEGFKVL 537

QY 541 YATDGFYATIPGKEPETIKKAKKELFKYNSKLPGLLELEYEGFYLRGFFVAKKRYAVI 600  
 DB 538 YATDGLGHATIPGADAEYVKKKXAFELNYPKLPGLLELEYEGFYVRGFFVTKKRYAVI 597  
 QY 601 DEEGRIITRGLGVVRDSEIAKEIQAKVLEBAILKEDSVKAVEIKVDVBEIAKYQVPL 660  
 DB 598 DEEGKIITRGLGVVRDSEIAKEIQAKVLEBAILKEDSVKAVEIKVDVBEIAKYQVPP 657  
 QY 661 EKLVIHEQITKDLSEYKAIGPHVAIAKRLAAGIKVRPGTIISYIVLRSGKISDRVILL 720  
 DB 658 EKLVIHEQITRELKDYKATGPHVAIAKRLAAGIKVRPGTIISYIVLRSGKISDRVILL 717  
 QY 721 SEYDPKHKYDPPYIENQVLPVAVURILEAFGRKEDLKQSSKQVGLDAWL 772  
 DB 718 DEEDPTKHYDADYIENQVLPVAVURILEAFGRKEDLKQSSKQVGLDAWL 769

RESULT 8  
 US-07-966-278-1  
 / Sequence 1, Application US/07966278  
 / Patent No. 5489523  
 / GENERAL INFORMATION:  
 / APPLICANT: Mathur, Eric A  
 / TITLE OF INVENTION: EXONUCLEASE-DEFICIENT THERMOSTABLE  
 / TITLE OF INVENTION: PYROCOCUS FURIOSUS DNA POLYMERASE I  
 / NUMBER OF SEQUENCES: 4  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Thomas Fitting  
 / STREET: 12526 High Bluff Road, Suite 300  
 / CITY: San Diego  
 / STATE: CA  
 / COUNTRY: USA  
 / ZIP: 92130  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patent in Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/07/966,278  
 / FILING DATE: 19921226  
 / CLASSIFICATION: 435  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER:  
 / FILING DATE:  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Fitting, Thomas  
 / REGISTRATION NUMBER: 34,163  
 / REFERENCE/DOCKET NUMBER: STG0133P  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 619-792-3680  
 / TELEFAX: 619-792-8477  
 / INFORMATION FOR SEQ ID NO: 1:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 775 amino acids  
 / TYPE: AMINO ACID  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / HYPOTHETICAL: NO  
 / ANTI-SENSE: NO  
 / US-07-966-278-1

Query Match 78.3%; Score 3151; DB 1; Length 775;  
 Best Local Similarity 74.8%; Pred. No. 7.4e-237;  
 Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

QY 1 MIFDYYITKDGKPIIRIFKENGEGFKIELDPHFQPIYIALLKDDSAIDEIKAKGERHG 60  
 DB 1 MILVDYITTEGKPIRVFKKDGFKIEYDFEPFIYALLRDSASIEIEKITAERHG 60  
 QY 61 KIVRVVDAVKKKKGLGRVVEWKLIFHPQVDPALRGKIREHPAVIDIYDIPFAKRY 120

Db 61 KIVRVDVKEVKFLGKPIVWKLYLHPQDVPTIREKVRHPAVVDIPEYDIPFARY 120  
Qy 121 LIDKGLIPMEGBELKLMADIFETFYHGEDEFGKGETIMISYADEBEARVITWKNIDLPY 180  
Db 121 LIDKGLIPMEGBELKLMADIFETFYHGEDEFGKGETIMISYADEBEARVITWKNIDLPY 180  
Qy 181 VDVVSNREMIKRFVQIVREKDPDVLITYNGDNFDPYLIKRAEKLGVTLLGRDKHEPE 240  
Db 181 VEVVSSREMIKFLRIIREKDPDIIVTYNGDSFDPYLAKEKRAEKLGIKLTIGRDS--E 238  
Qy 241 PKIHRMGDSFAVEIKGRIHFDLPVVRRTINLPTTYLEAVYEAVLGKTSKLGAEIAAI 300  
Db 239 PKMQRIGDMTAVEVKGRHFDLYHVITRTINLPTTYLEAVYEAIFGKPKKVVADIAKA 298  
Qy 301 WETEESMKLAQYSMEDARATYELGKEFPFMEAEALAKLGOSVWDVSRSTGNLVEWYLL 360  
Db 299 WESGENLVRVAKYSMEDAKATYELGKEFLPMETQLSRLVGQPLWDVSRSTGNLVEWFL 358  
Qy 361 RVAYERNELAPNKPDEEYRRRTTTLGGYVKEPERGLWENITLDFCLYPSIITWN 420  
Db 359 RKAYERNEVAPNKPSEEEYQRRRESYTGFGVKEPEKGLWENIVLDFRALYPSIITHN 418  
Qy 421 VSPDTLRECKQNDVAPIVGYKCDPFGFIPSIILGELITMRQEIKKOKKATIDIEKK 480  
Db 419 VSPDTLNLGCKNYDIAPOVGHKFCDDIPGFIPSLGLHLEERQKIKTKVKTQDPIEKI 478  
Qy 481 MLDYORAVKLVHANSYGYMGYPKARWYSKECAESVTANGRHYEMTIEIEBKFGFKVL 540  
Db 479 LLDYORQAKILLANSFYGYGAKARWYCKEACSVTANGRKYIELVWKELEKEFGFKVL 538  
Qy 541 YADTDGYATIPGKEPETIKKAKPEFLKINSKLPGLLEVEYEGFYLRGFFVAKKRYAVI 600  
Db 539 YIDTDGLYATIPGSESEIKKALEFVKVINSKLPGLLEVEYEGFYLRGFFVTKRYAVI 598  
Qy 601 DEGRITTRGLEVWRDSEIAKETQAKVLEBALKEDSVKEAVEIVKDVVEIAKYQVPL 660  
Db 599 DEBKVITRGLVWRDSEIAKETQARVLETLRHGDVEEAIRIVKEVIQKLANYEIPP 658  
Qy 661 EKLVIHQITKOLSEYKAIQPHVAIAKRAAAGIKVRPGTIISYVLRGSGKISDRVILL 720  
Db 659 EKLAIEQITRPLEYKAIQPHVAVAKLAAGKVKIPGMWIGYVLRGDSIPSNRAIIA 718  
Qy 721 SEYDPKHKYDDPYIENQVLFVLAILEAFGRKEDLYQSSKQVGLDAWL 772  
Db 719 BEYDPKHKYDABYIENQVLFVLAILEFGYRKEDLYQKTRQVGLTSL 770

## RESULT 9

US-08-424-921-1  
; Sequence 1, Application US/08424921  
; Patent No. 5545552  
; GENERAL INFORMATION:  
; APPLICANT: Mathur, Eric A  
; TITLE OF INVENTION: PURIFIED THERMOSTABLE PYROCOCUS  
; TITLE OF INVENTION: FURIOSUS DNA POLYMERASE I  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bingham & Pitting  
; STREET: 12526 High Bluff Road, Suite 300  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92130  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424,921  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/803,627  
; FILING DATE: 02-DEC-1991  
; APPLICATION NUMBER: US 07/620,568  
; FILING DATE: 03-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/657,073  
; FILING DATE: 19-FEB-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/776,552  
; FILING DATE: 15-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: STG0100P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-792-3680  
; TELEFAX: 619-792-8477  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 775 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; US-08-424-921-1

Query Match 78.3%; Score 3151; DB 1: Length 775;  
Best Local Similarity 74.6%; Pred. No. 7.4e-937;

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

Qy 1 MIFDTITDQGPITIRIFKENGEGFKIEDPHFPQYIYALLKDDSAIDEIKAGSRHG 60  
Db 1 MLDVDYITBEGFVIRLFKENGKFKIEHRTFRPIYALLRDDSDKIEBVKKITGRHG 60  
Qy 61 KIVRVDVAVKVKKFLGRDVEVWKLIFEPHPQDVPALRGKIREHPAVTDIYEYDIPAKRY 120  
Db 61 KIVRVDVAVKVKKFLGRDVEVWKLIFEPHPQDVPALRGKIREHPAVTDIYEYDIPAKRY 120  
Qy 121 LIDKGLIPMEGBELKLMADIFETFYHGEDEFGKGETIMISYADEBEARVITWKNIDLPY 180  
Db 121 LIDKGLIPMEGBELKLMADIFETFYHGEDEFGKGETIMISYADEBEARVITWKNIDLPY 180  
Qy 181 VDVVSNREMIKRFVQIVREKDPDVLITYNGDNFDPYLIKRAEKLGVTLLGRDKHEPE 240  
Db 181 VEVVSSREMIKFLRIIREKDPDIIVTYNGDSFDPYLAKEKRAEKLGIKLTIGRDS--E 238  
Qy 241 PKIHRMGDSFAVEIKGRIHFDLPVVRRTINLPTTYLEAVYEAVLGKTSKLGAEIAAI 300  
Db 239 PKMQRIGDMTAVEVKGRHFDLYHVITRTINLPTTYLEAVYEAIFGKPKKVVADIAKA 298  
Qy 301 WETEESMKLAQYSMEDARATYELGKEFPFMEAEALAKLGOSVWDVSRSTGNLVEWYLL 360  
Db 299 WESGENLVRVAKYSMEDAKATYELGKEFLPMETQLSRLVGQPLWDVSRSTGNLVEWFL 358  
Qy 361 RVAYERNELAPNKPDEEYRRRTTTLGGYVKEPERGLWENITLDFCLYPSIITWN 420  
Db 359 RKAYERNEVAPNKPSEEEYQRRRESYTGFGVKEPEKGLWENIVLDFRALYPSIITHN 418  
Qy 421 VSPDTLRECKQNDVAPIVGYKCDPFGFIPSIILGELITMRQEIKKOKKATIDIEKK 480  
Db 419 VSPDTLNLGCKNYDIAPOVGHKFCDDIPGFIPSLGLHLEERQKIKTKVKTQDPIEKI 478  
Qy 481 MLDYORAVKLVHANSYGYMGYPKARWYSKECAESVTANGRHYEMTIEIEBKFGFKVL 540  
Db 479 LLDYORQAKILLANSFYGYGAKARWYCKEACSVTANGRKYIELVWKELEKEFGFKVL 538  
Qy 541 YADTDGYATIPGKEPETIKKAKPEFLKINSKLPGLLEVEYEGFYLRGFFVAKKRYAVI 600  
Db 539 YIDTDGLYATIPGSESEIKKALEFVKVINSKLPGLLEVEYEGFYLRGFFVTKRYAVI 598  
Qy 601 DEGRITTRGLEVWRDSEIAKETQAKVLEBALKEDSVKEAVEIVKDVVEIAKYQVPL 660  
Db 719 BEYDPKHKYDABYIENQVLFVLAILEFGYRKEDLYQKTRQVGLTSL 770



Db 599 DEEGKVTIRGLIIVRRDWSIAKETQARVLETILKHGDVEAVRIYKEVIQKLANYEIPP 658  
Qy 661 EKLVIHQITKDLSEYKAIGHVALAKLAAGIKVRRPGTIIYIVLRSGKTSRVL 720  
Db 659 EKLAIEQITRPLHEYKAIGHVALAKLAAGIKVRRPGTIIYIVLRSGKTSRVL 718  
Qy 721 SEYDPKHKYDPPDYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAML 772  
Db 719 EYDPPKHKYDAEYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAML 770

RESULT 10  
US-08-556-355A-1  
; Sequence 1, Application US/08556355A  
; Patent No. 5966395  
; GENERAL INFORMATION:  
; APPLICANT: MATHUR, Eric J.  
; TITLE OF INVENTION: Purified Thermostable Pyrococcus  
; TITLE OF INVENTION: furiosus DNA Polymerase I  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESS: Dunner, L.L.P.  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/556.355A  
; FILING DATE: 13-NOV-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/424,921  
; FILING DATE: 19-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/803,627  
; FILING DATE: 02-DEC-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/779,846  
; FILING DATE: 21-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/657,073  
; FILING DATE: 19-FEB-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/620,568  
; FILING DATE: 03-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barker, M. Paul  
; REGISTRATION NUMBER: 32,013  
; REFERENCE/DOCKET NUMBER: 04121.0004-02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO. 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 775 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-556-355A-1

Query Match 78.3%; Score 3151; DB 2; Length 775;  
Best Local Similarity 74.6%; Pred. NO. 7.4e-237;  
Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

Qy 1 MIFDVTYTKDGKPIRIFKENGEPKIELDPHFQPIYALKKDDSAIDELKAIKGERHG 60

Db 1 MILDVDYITEGKPVIRLFKENGKFKLEHDTTRPYIYALLRDDSKIEEVKKTGERHG 60  
Qy 61 KIVRVVDVAVKVKFLGRDVEVWKLIFEPHODVPALRGKIREHPAVDIYDIYDIPPAKRY 120  
Db 61 KIVRIVDVAVKVKFLGRDVEVWKLIFEPHODVPALRGKIREHPAVDIYDIYDIPPAKRY 120  
Qy 121 LIDKGLIPMEGDEELKLMFADIEYFHEGDFGKEIIMISYADEEAEARVITWKNIDLPY 180  
Db 121 LIDKGLIPMEGDEELKLMFADIEYFHEGDFGKEIIMISYADEEAEARVITWKNIDLPY 180  
Qy 181 VDVSNREMIKRFVQIVREKDPVLTINYNDNFDLPYLIRAKGLGVTLGLGRDKEHPE 240  
Db 181 VDVSNREMIKRFVQIVREKDPVLTINYNDNFDLPYLIRAKGLGVTLGLGRDKEHPE 240  
Qy 241 PKIHRMGDSFAVEIKGRHFDLPFVVRTINLPVTLAEAVYEAIVLGTGSKLGAEEAAI 300  
Db 241 PKIHRMGDSFAVEIKGRHFDLPFVVRTINLPVTLAEAVYEAIVLGTGSKLGAEEAAI 300  
Qy 298 PKQRIQDMTAVEVKGRIHFDLYHVIPTINLPVTLAEAVYEAIVLGTGSKLGAEEAAI 360  
Db 298 PKQRIQDMTAVEVKGRIHFDLYHVIPTINLPVTLAEAVYEAIVLGTGSKLGAEEAAI 360  
Qy 301 METESMKLAQYSMEDARATYELGKEFPFMEALAKIIGQSVMDVSRSSGTNLVWELL 360  
Db 301 METESMKLAQYSMEDARATYELGKEFPFMEALAKIIGQSVMDVSRSSGTNLVWELL 360  
Qy 358 MESGENLERVAKYSMEDAKATYELGKEFPFMEALAKIIGQSVMDVSRSSGTNLVWELL 358  
Db 358 MESGENLERVAKYSMEDAKATYELGKEFPFMEALAKIIGQSVMDVSRSSGTNLVWELL 358  
Qy 361 RVAVERNELAPNKDEEYRRLRTYLGGVVKEPERGLWENITVLDPRCLYPSIIVTHN 420  
Db 361 RVAVERNELAPNKDEEYRRLRTYLGGVVKEPERGLWENITVLDPRCLYPSIIVTHN 420  
Qy 359 RKAYERNEVAPNKPSESEYQRLRESYTGQFKPEKGLWENIYVLDPRALYPSIIVTHN 418  
Db 359 RKAYERNEVAPNKPSESEYQRLRESYTGQFKPEKGLWENIYVLDPRALYPSIIVTHN 418  
Qy 421 VSPDTLREGCKNYDVAPIVGYKFCOKPFGPIPSILGELITWRQEIKKKATIDPIBK 480  
Db 421 VSPDTLREGCKNYDVAPIVGYKFCOKPFGPIPSILGELITWRQEIKKKATIDPIBK 480  
Qy 419 VSPDTLREGCKNYDVAPIVGYKFCOKPFGPIPSILGELITWRQEIKKKATIDPIBK 478  
Db 419 VSPDTLREGCKNYDVAPIVGYKFCOKPFGPIPSILGELITWRQEIKKKATIDPIBK 478  
Qy 481 MLDYRQRAVKLHANSYGYMGYPKARYSKCARSVTAGRWYIEMTKETEEKFGKVL 540  
Db 481 MLDYRQRAVKLHANSYGYMGYPKARYSKCARSVTAGRWYIEMTKETEEKFGKVL 540  
Qy 479 LLDYRQRAVKLHANSYGYMGYPKARYSKCARSVTAGRWYIEMTKETEEKFGKVL 538  
Db 479 LLDYRQRAVKLHANSYGYMGYPKARYSKCARSVTAGRWYIEMTKETEEKFGKVL 538  
Qy 541 YADTDGTYATIPGKEPETIKKAKKFLKYINSKPLGLELEYEGFYLGGFFVKKRYAVI 600  
Db 541 YADTDGTYATIPGKEPETIKKAKKFLKYINSKPLGLELEYEGFYLGGFFVKKRYAVI 600  
Qy 539 YIDTDGLYATIPGGESEIEKKKALEFVKYINSKPLGLELEYEGFYLGGFFVKKRYAVI 598  
Db 539 YIDTDGLYATIPGGESEIEKKKALEFVKYINSKPLGLELEYEGFYLGGFFVKKRYAVI 598  
Qy 601 DEEGITRGLIIVRRDWSIAKETQARVLETILKHGDVEAVRIYKEVIQKLANYEIPP 660  
Db 601 DEEGITRGLIIVRRDWSIAKETQARVLETILKHGDVEAVRIYKEVIQKLANYEIPP 660  
Qy 599 DEEGKVTIRGLIIVRRDWSIAKETQARVLETILKHGDVEAVRIYKEVIQKLANYEIPP 658  
Db 599 DEEGKVTIRGLIIVRRDWSIAKETQARVLETILKHGDVEAVRIYKEVIQKLANYEIPP 658  
Qy 661 EKLVIHQITKDLSEYKAIGHVALAKLAAGIKVRRPGTIIYIVLRSGKTSRVL 720  
Db 661 EKLVIHQITKDLSEYKAIGHVALAKLAAGIKVRRPGTIIYIVLRSGKTSRVL 720  
Qy 721 SEYDPKHKYDPPDYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAML 772  
Db 721 SEYDPKHKYDPPDYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAML 772  
Qy 719 EYDPPKHKYDAEYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAML 770  
Db 719 EYDPPKHKYDAEYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAML 770

RESULT 11  
US-07-803-627A-1  
; Sequence 1, Application US/07803627A  
; Patent No. 5948663  
; GENERAL INFORMATION:  
; APPLICANT: MATHUR, Eric J.  
; TITLE OF INVENTION: Purified Thermostable Pyrococcus  
; TITLE OF INVENTION: furiosus DNA Polymerase I  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESS: Dunner, L.L.P.  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

601 DEGRITRGLVVRDSEIATKTOAKVLEAILKEDSVEKAVIIVKDVVEEIAKYQVPL 660  
659 DEGRITRGLVVRDSEIATKTOAKVLEAILKEDSVEKAVIIVKDVVEEIAKYQVPL 658  
661 EKLVIHEQITKDLSEYKAIGHVAIAKRLAAKGIKVRPGTIIISVILRGSGKISDRVILL 720  
659 EKLAIYEQITEPLEHXYKAIGHVAIAKRLAAKGIKVRPGTIIISVILRGSGKISDRVILL 718  
721 SEYDPKHKHDPDYIENQVLPVLRILEAFGVYKEDLKYQSSKQVGLDAML 772  
719 EYDPKHKHDAEYIENQVLPVLRILEAFGVYKEDLKYQSSKQVGLDAML 770

RESULT 12  
US-09-244-889A-1  
Sequence 1, Application US/09244889A  
Patent No. 6489150  
GENERAL INFORMATION:  
APPLICANT: MATHUR, Eric J.  
TITLE OF INVENTION: Purified Thermostable Pyrococcus  
FURIOSUS DNA POLYMERASE I  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/244,889A  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 09/135,064  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/803,627  
FILING DATE: 02-DEC-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/776,552  
FILING DATE: 15-OCT-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,073  
FILING DATE: 19-FEB-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/620,568  
FILING DATE: 03-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Barker, M. Paul  
REGISTRATION NUMBER: 32,013  
REFERENCE/DOCKET NUMBER: 04121.0004-01  
TELEPHONE: (202)408-4000  
TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-244-889A-1

Query Match 78.3%; Score 3151; DB 4; Length 775;  
Best Local Similarity 74.6%; Pred. No. 7.4e-237;  
Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/803,627A  
FILING DATE: 02-DEC-1991  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/776,552  
FILING DATE: 14-OCT-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,073  
FILING DATE: 19-FEB-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/620,568  
FILING DATE: 03-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Barker, M. Paul  
REGISTRATION NUMBER: 32,013  
REFERENCE/DOCKET NUMBER: 04121.0004-00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)408-4000  
TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 775 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-803-627A-1

Query Match 78.3%; Score 3151; DB 2; Length 775;  
Best Local Similarity 74.6%; Pred. No. 7.4e-237;  
Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

QY 1 MIFDTYITKDGKPIIRIKPKNGKFKIELDHPQVPIYVALLKDSATDEIKAIKGERHG 60  
DB 1 MILDVDYITEGKPVIRLFPKNGKFKIEHDTFRPIYVALLRDSKIEEVKKTGERHG 60  
QY 61 KIVRVVDVAVKVKKGLGRDVEVVKLIFHPQVPAIRGKIREHPAVIDIYEDIPFAKRY 120  
DB 61 KIVRIVDVEKVKKGLGKPIITWKLYLEHPQVPIREKVRPAVIDIYEDIPFAKRY 120  
QY 121 LIDKGLIPMEGDEELKMAFDIETTHGDEFGKGEIIMISVADDEEAEARVITWKNDLPY 180  
DB 121 LIDKGLIPMEGDEELKMAFDIETTHGDEFGKGEIIMISVADDEEAEARVITWKNDLPY 180  
QY 181 VDVVNSNEREMIKRFQVIREKDPDVLITVNGNDFLVLKRAEKLGVTLALGRDKEHPE 240  
DB 181 VEVVSSEREMIKRFIIRKDPDIIIVNGDSFDFPVLAKRAEKLGVTLALGRDKEHPE 240  
QY 241 PKIHRMGDSFAVEIKRIFHDFLPVVRRTINLPVTVLEAVVEAVLGTCKSKLGAEEIAAI 300  
DB 239 PKVQIGDMTAVEVKGRIHDFLVHVTITNLPVTVLEAVVEAVLGTCKSKLGAEEIAAI 298  
QY 301 WETESMKKLAQYSMEDARATVELCKEFPWAEALAKLIGQSVWDSSTGNLVEWVLL 360  
DB 299 WESGENLERVAKYSHEDAKATVELCKEFLPMETIQLSRVVGQPLMDVSRSSSTGNLVEWVLL 358  
QY 361 RVAYERNELAPKPDDEEYRRLRTTYLGGYKPEPGLWENITVLDPRCLYPSIIVTHN 420  
DB 359 RXAYERNEVAPKPSSEYQRRLESYTGKFKPEKGLWENIVYLDPRALYPSIIVTHN 418  
QY 421 VSPDTLEREGCKNDYVPIVGVKCKDPGRTIPSLITWREIKKMAKATIDPIEKK 480  
DB 419 VSPDTLNEGCKNDYVPIVGVKCKDPGRTIPSLITWREIKKMAKATIDPIEKK 478  
QY 481 MLDYTSQRAVKLHANSYGYMGYPKARWYSKECAESVTAWGRHYEMTIKEEKFPGKVL 540  
DB 479 LLDYRQKAIKLLANSFYGYGAKARWYSKECAESVTAWGRHYEMTIKEEKFPGKVL 538  
QY 541 YADTDGFYATPGEKPEIKKAKFVKYINSKLPGLLELEYEGFYLRGFFVAKRYAVI 600  
DB 539 YIDTDGLYATPGESEIEKKAEFVKYINSKLPGLLELEYEGFYLRGFFVAKRYAVI 598

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;  
QY 1 MIFDVTYTKGPIIRIFPKKENGEBKIELDPHFQVIVALLKDDSAIDEIKAIKGERHG 60  
DB 1 MILDVDYITEGKPVIRLFKENGKFKIEHDTFRFYIYALLRDDSKIEEVKKITGERHG 60  
QY 61 KIVRVDAVKVKKFLGRDVEVWKLIFHPQDVPAIRKIREHPAVIDIYEDIPPAKRY 120  
DB 61 KIVRIVDVEKVKFLGKPIITVWKLYLEHPQDVPTIREKVRHPAVVDIYEDIPPAKRY 120  
QY 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKEIIMISYADEEAEARVITWKNIDLPY 180  
DB 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKEIIMISYADENEAKVITWKNIDLPY 180  
QY 181 VDVVSNEREMIKRFVQIVREKDPDVLITYNGDNFDPYLIKRAEKLGVTLGLGRDKEHPE 240  
DB 181 VEVVSSEEMIKRFLRIIREKDPDIIVTYNGDSFDPYLAKEKLGKLTIGRDS--E 238  
QY 241 PKIHEMGDSFAVEIKGRIFHDLFPVVRTINLPYTLLEAVYEAVLGKTKSGABEIAAI 300  
DB 239 PKMQRIGDMTAVEVGRHLHFDLHVITRTINLPYTLLEAVYEAIFGKPKYVADDEIAKA 298  
QY 301 WETESMKKLAQYSMEDARATYELGKEFPFMAELAKLIGQSVMDVSRSTGNLVEWYLL 360  
DB 299 WESGENLERVAKYSMEDAKATYELGKEFLPMEIQLSRLVGQPLMDVSRSTGNLVEWYLL 358  
QY 361 RVAYERNELAPNKPDEEYRRLRTYLGYYVKEPERGLWENITYLDPRCLYPSIIVTHN 420  
DB 359 RKAYERNEVAENKPSSEYQRLRESYTGFGVKEPEKGLWENIYLDPRALYPSIIVTHN 418  
QY 421 VSPDTLREGECKNYDVAPIVGYKFCCKDPGFPFIPSLGELITMRQIEKKKATIDPIEK 480  
DB 419 VSPDTLNEGCKNYDIAPOVGHKFCCKDIPGFIPLGLHLEERQKIKTKMKEQDPIEKI 478  
QY 481 MLDYRQRAVKLHANSYGYMGYPKARWYKCAESVTANGRHVITWTKIEEKEGFKVL 540  
DB 479 LLDYRQRAIKLANSFYGYGYKARWYKCAESVTANGRKYIELVWKELEKEGFKVL 538  
QY 541 YADTDGFIATIPGKPEIKKAEFLKYINSKLPGLLELEYEGFYLRGFFVAKRYAVI 600  
DB 539 YIDTDGLYATIPGSESEIKKALEFVKYINSKLPGLLELEYEGFYKGEFFVTKRYAVI 598  
QY 601 DEGRITRGLVVRDWESEIAKTQAKVLEAILKEDSVEKAVEIVKDVVEEIAKQVPL 660  
DB 599 DEEGKVIPTRGLEIVRDWESEIAKTQARVLETLKHGDVEEAVRIVKEVIQKLANVEIPP 658  
QY 661 EKLVIHEOITLDSYKALGPHVAIAKLAAGIKVRPGTIIISIVILRGSKISDRVILL 720  
DB 659 EKLAIEQITPLHEYKALGPHVAVAKLAAGVYKIPGVMIGVIVILRGDGPISNRALLA 718  
QY 721 SEYDPKHKYDPPDYIENQVLPVLRILEAFGYRKEDLKYOSSKQVGLDAWL 772  
DB 719 BEYDPKHKYDAEYIENQVLPVLRILEGFGYRKEDLRYQKTRQVGLTSLW 770

## RESULT 13

US-08-688-649-37

Sequence 37, Application US/08688649

Patent No. 5827716

GENERAL INFORMATION:

APPLICANT: MAMONE, JOSEPH A.

TITLE OF INVENTION: MODIFIED POL-II TYPE DNA

TITLE OF INVENTION: POLYMERASES

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon &amp; Lyon

STREET: 633 West Fifth Street

STREET: Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/688,649  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 220/281  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 776 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-688-649-37

Query Match 78.3%; Score 3151; DB 2; Length 776;

Best Local Similarity 74.6%; Pred. No. 7.4e-237;

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

QY 1 MIFDVTYTKGPIIRIFPKKENGEBKIELDPHFQVIVALLKDDSAIDEIKAIKGERHG 60  
DB 1 MILDVDYITEGKPVIRLFKENGKFKIEHDTFRFYIYALLRDDSKIEEVKKITGERHG 60  
QY 61 KIVRVDAVKVKKFLGRDVEVWKLIFHPQDVPAIRKIREHPAVIDIYEDIPPAKRY 120  
DB 61 KIVRIVDVEKVKFLGKPIITVWKLYLEHPQDVPTIREKVRHPAVVDIYEDIPPAKRY 120  
QY 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKEIIMISYADEEAEARVITWKNIDLPY 180  
DB 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKEIIMISYADENEAKVITWKNIDLPY 180  
QY 181 VDVVSNEREMIKRFVQIVREKDPDVLITYNGDNFDPYLIKRAEKLGVTLGLGRDKEHPE 240  
DB 181 VEVVSSEEMIKRFLRIIREKDPDIIVTYNGDSFDPYLAKEKLGKLTIGRDS--E 238  
QY 241 PKIHEMGDSFAVEIKGRIFHDLFPVVRTINLPYTLLEAVYEAVLGKTKSGABEIAAI 300  
DB 239 PKMQRIGDMTAVEVGRHLHFDLHVITRTINLPYTLLEAVYEAIFGKPKYVADDEIAKA 298  
QY 301 WETESMKKLAQYSMEDARATYELGKEFPFMAELAKLIGQSVMDVSRSTGNLVEWYLL 360  
DB 299 WESGENLERVAKYSMEDAKATYELGKEFLPMEIQLSRLVGQPLMDVSRSTGNLVEWYLL 358  
QY 361 RVAYERNELAPNKPDEEYRRLRTYLGYYVKEPERGLWENITYLDPRCLYPSIIVTHN 420  
DB 359 RKAYERNEVAENKPSSEYQRLRESYTGFGVKEPEKGLWENIYLDPRALYPSIIVTHN 418  
QY 421 VSPDTLREGECKNYDVAPIVGYKFCCKDPGFPFIPSLGELITMRQIEKKKATIDPIEK 480  
DB 419 VSPDTLNEGCKNYDIAPOVGHKFCCKDIPGFIPLGLHLEERQKIKTKMKEQDPIEKI 478  
QY 481 MLDYRQRAVKLHANSYGYMGYPKARWYKCAESVTANGRHVITWTKIEEKEGFKVL 540  
DB 479 LLDYRQRAIKLANSFYGYGYKARWYKCAESVTANGRKYIELVWKELEKEGFKVL 538  
QY 541 YADTDGFIATIPGKPEIKKAEFLKYINSKLPGLLELEYEGFYLRGFFVAKRYAVI 600  
DB 539 YIDTDGLYATIPGSESEIKKALEFVKYINSKLPGLLELEYEGFYKGEFFVTKRYAVI 598



Db 1 MILDQVITBEGKPVIRVEKKDGEFKI EYDRFEFPIYVIALLRDSDSAIBEIEKI TAERHG 60  
Qy 61 KIYRVVDVAVKVKKFLGRDVEVWKL IFEHPQDVPAIR - GKIREHPAVIDIYDYDIPFAKR 119  
Db 61 KWKVKKAEKVKKKFLGRSEVWVLYFTHPQDVPAIRPKIRKHPAVIDIYDYDIPFAKR 120  
Qy 120 YLIDKGLIPMEGDEBELKMAFDIETFYHGEDEFGKGEIIMISYADEEERARVITWKIDLP 179  
Db 121 YLIDKGLIPMEGDEBELKMSFDIETLYHEGEFEFGTGPILMISYADESEARVITWKIDLP 180  
Qy 180 YVDVVSNEREMIKFVQIVREKDDULITNGDNFDLPYL IKAELKGVTL LGRDKEHP 239  
Db 181 YVDVVSNEKEMIKRFLKVKVEKDPDLITTYDGNDFPAY LKRCCEKLGVSFTLGRDGS-- 238  
Qy 240 EPKIRMGDSFAVEIKRIHFDLPFPVVRRTINLPTTYTLEAVYEA VLGKTKSLGAEI AA 299  
Db 239 EPKIQRMGDRFAVEVKGRHIFDLYPAIRRTINLPTTYTLEAVYEA VLGKTKSLGAEI AA 298  
Qy 300 IWETEESMKLAQVSMEDADAATYELGKEFPMAELAKLIGQSVWDVSRSSSTGNLVEWYL 359  
Db 299 AWETGELEGVARYSMEDARVYELGREFFPMEQAQLSRLLIGQGLWDVSRSSSTGNLVEWEL 358  
Qy 360 LRVAVERNELAPNKPDEEYRRRLRTTYLGGYVKEPERGLWENIT YLDRCLYPSIIITH 419  
Db 359 LRKAYERNELAPNKPDERELARR - RGGYAGYVKEPERGLWENIT YLDRCLYPSIIITH 417  
Qy 420 NVSPDTLREGCKNYDVAPIGVYKFCXDPFGFIPSLGELITWRQSIKKMKATIDPIEK 479  
Db 418 NVSPDTLNRREGCKSYDAPQVGHKFCXDPFGFIPSLGELITWRQSIKKMKATIDPIEK 477  
Qy 480 KMLDYRORAVKLHANSYGYMGYPKARWYSKECAESVTAWGRHYEWTI KEIEEKEGFKV 539  
Db 478 KLLDYRQRAIKILANSFYGYGYABARWYCKECAESVTANGREYIEMVIRELEKEGFKD 537  
Qy 540 LYADTDGFYATIPGEKETIKKAKPEFLKYINSKLPGLLELEYEGFYLRGFFVAKKRYAV 599  
Db 538 LYADTDGLHATIPGADRETIVKKDLEFLNINPKLPGLLELEYEGFYLRGFFVAKKRYAV 597  
Qy 600 IDEEGRIITRGLVVRDWSIELAKETOAKVLEAILKEDSVKAVEIVKQVVEI AKYQVP 659  
Db 598 IDEEGKITRGLVVRDWSIELAKETILARVLEAILRHDVEEAVRIKRETEKLSKIEVP 657  
Qy 660 LEKLVIEHQITKDLSEYKAIGPHVAIAKRLAAKGIKVRPGTIIISYIVLRGSGKISDRVIL 719  
Db 658 PEKLVITEQITRELKDYKATGPHVAIAKRLAAKGIKVRPGTIIISYIVLRGSGKISDRVIL 717  
Qy 720 LSEYDPKHKYDPPDYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWL 772  
Db 718 PDEFDPKHYDADYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWL 770

Search completed: February 9, 2004, 11:53:46  
Job time : 34'secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2004, 11:46:55 ; Search time 40 Seconds  
(without alignments)  
1860.864 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026

Sequence: 1 MIFDTDYITKDGKPIIRIFK.....KEDLKYQSSKQVGLDAMLKK 774

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query March	Length DB	ID	Description
1	3356	83.4	1702	2 S42459	DNA-directed DNA p
2	3204.5	79.6	775	2 S67920	DNA-directed DNA p
3	3164.5	78.6	771	2 S75023	DNA polymerase I p
4	3151	78.3	775	2 S35543	DNA-directed DNA p
5	2934.5	72.9	1312	2 S28593	DNA-directed DNA p
6	2886	71.7	1235	2 C1210	probable DNA-direc
7	2767.5	68.7	1670	2 S71551	DNA-directed DNA p
8	1301	32.3	781	2 A69312	DNA polymerase B1
9	1291	32.1	1634	2 E64410	DNA-directed DNA p
10	1209.5	30.0	586	2 C69028	DNA-dependent DNA p
11	1204.5	29.9	781	2 UC7382	DNA-directed DNA p
12	1190	29.6	803	2 B56277	DNA-directed DNA p
13	1138	28.3	784	2 E2515	probable DNA-direc
14	697.5	17.2	1088	2 T05731	DNA-directed DNA p
15	676.5	16.8	764	2 S75407	probable DNA-direc
16	659.5	16.4	1038	1 JC5757	DNA-directed DNA p
17	657.5	16.3	1038	2 T18222	DNA polymerase del
18	657.5	16.3	1086	2 T42242	DNA-directed DNA p
19	649.5	16.1	901	2 T43266	DNA-directed DNA p
20	626	15.5	901	2 E43210	DNA polymerase B1
21	621.5	15.4	1097	1 RNEYJ3	DNA-directed DNA p
22	619.5	15.4	882	2 S23019	DNA-directed DNA p
23	619.5	15.4	1084	1 S19661	DNA-directed DNA p
24	616.5	15.3	872	2 JC7380	DNA-directed DNA p
25	616.5	15.3	882	2 F90201	DNA polymerase I
26	616.5	15.3	1095	1 S40243	DNA-directed DNA p
27	614.5	15.3	1094	2 S22573	DNA-directed DNA p
28	611	15.2	1107	1 A41618	DNA-directed DNA p
29	609	15.1	1513	2 T28158	probable DNA-direc

30	605.5	15.0	1106	1 A39299	DNA-directed DNA p
31	587	14.6	1492	2 T18560	DNA-directed DNA p
32	584.5	14.5	879	2 A56277	DNA-directed DNA p
33	583.5	14.5	1462	1 DJHUAC	hypothetical prote
34	581	14.4	1081	2 T20698	DNA-directed DNA p
35	578.5	14.4	875	2 JCS186	probable DNA-direc
36	570.5	14.2	959	2 F72763	DNA-directed DNA p
37	570	14.2	1465	2 S45628	DNA-dependent DNA p
38	559.5	13.9	223	2 E69125	DNA-directed DNA p
39	548.5	13.6	1339	1 S20052	DNA-directed DNA p
40	527.5	13.1	1488	2 S58250	DNA-directed DNA p
41	522.5	13.0	1505	2 S28079	DNA-directed DNA p
42	516.5	12.8	844	2 T31321	DNA-directed DNA p
43	516	12.8	1405	1 DJZPA	hypothetical prote
44	504.5	12.5	2154	2 A84669	DNA-directed DNA p
45	500.5	12.4	913	2 T17675	DNA-directed DNA p

#### ALIGNMENTS

##### RESULT 1

S42459

N: Containing DNA polymerase (EC 2.7.7.7) Vent, interin containing precursor - Thermococcc  
C: Species: Thermococcus litoralis  
C: Date: 31-Dec-1993 #sequence revision 02-Aug-1994 #text\_change 18-Jun-1999  
C: Accession: S42459; S42451; S42450; S42458

R: Perler, F.B.; Comb, D.G.; Jack, W.E.; Moran, L.S.; Qiang, B.; Kucera, R.B.; Benner, J  
submitted to the EMBL Data Library, September 1992

A: Reference number: S42458

A: Accession: S42459

A: Molecule type: DNA

A: Residues: 1-1702 <PER1>

A: Cross-references: EMBL:M74198; NID:G154685; PID:AAA72100.1; PID:G154686

R: Perler, F.B.; Comb, D.G.; Jack, W.E.; Moran, L.S.; Qiang, B.; Kucera, R.B.; Benner, J  
Proc. Natl. Acad. Sci. U.S.A. 89, 5577-5581, 1992

A: Title: Intervening sequences in an Archaea DNA polymerase gene.

A: Reference number: S42450; MUID:92302285; PMID:1608969

A: Accession: S42451

A: Molecule type: DNA

A: Residues: 181-222;387-425;452-476;483-524;1021-1062;1076-1099;1466-1489;1533-1547 <PE  
A: Cross-references: EMBL:M74198

R: Hodges, R.A.; Perler, F.B.; Noren, C.J.; Jack, W.E.

Nucleic Acids Res. 20, 6153-6157, 1992

A: Title: Protein splicing removes intervening sequences in an archaea DNA polymerase.

A: Reference number: S40786; MUID:93117083; PMID:1475179

A: Contents: annotation

C: Function: <VENT>

A: Description: nucleotidyltransferase

A: Note: DNA-directed DNA polymerase Vent

C: Function: <END1>

A: Description: endonuclease; hydrolase

A: Note: DNA endonuclease PI-Tlii

C: Function: <END2>

A: Description: endonuclease; hydrolase

A: Note: DNA endonuclease PI-Tlii

C: Superfamily: Thermococcus litoralis DNA-directed DNA polymerase Vent

C: Keywords: DNA binding; endonuclease; hydrolase; nucleotidyltransferase; protein splic

F: 1-494/1033-1081.1472-1702/Product; DNA-directed DNA polymerase Vent #status predicted

F: 1-494/Domain: DNA-directed DNA polymerase Vent extein 1 #status predicted <XT1>

F: 495-1032/Product; DNA endonuclease PI-II (pol Vent intein 1) #status predicted <MAT2>

F: 1033-1081/Domain: DNA-directed DNA polymerase Vent extein 2 #status predicted <XT2>

F: 1082-1471/Product; DNA endonuclease PI-I (pol Vent intein 2) #status predicted <MAT3>

F: 1472-1702/Domain: DNA-directed DNA polymerase Vent extein 3 #status predicted <XT3>

F: 494-1033/Cross-link: peptide (Asn-Ser) #status predicted

F: 1081-1472/Cross-link: peptide (Asp-Thr) #status predicted

Query Match 83.4%; Score 3356; DB 2; Length 1702;

Best Local Similarity 42.7%; Pred. No. 2e-164;

Matches 726; Conservative 33; Mismatches 15; Indels 928; Gaps 2;

QY 1 MIFDTDYITKDGKPIIRIFKENGFKIELDPHFQPIYALLKDDSAIDKAIKGRHG 60

Db 1 MILDVTITKGGKPIIRIFKENGEPKIELDPHPQPIYALLKDDSAIEBIKAIGERHG 60  
QY 61 KIVRVDAVKKKFELGRDVEVWKLIFEHQDVPALRGKIREHPAVDIDYEYDIPAKY 120  
Db 61 KTVRVDAVKKKFELGRDVEVWKLIFEHQDVPAMRGKIREHPAVVDIYEYDIPAKY 120  
QY 121 LIDKGLIPMEGDEBELKMAFDIETFYHEGDEFKGEIIMISYADEEAEARVITWKNIDL 180  
Db 121 LIDKGLIPMEGDEBELKMAFDIETFYHEGDEFKGEIIMISYADEEAEARVITWKNIDL 180  
QY 181 VDVVSNREMIKRVQVIREKDPVLTITNGDNFDLPYLKRAEKLGVTLILGRDKEHPE 240  
Db 181 VDVVSNREMIKRVQVIREKDPVLTITNGDNFDLPYLKRAEKLGVTLILGRDKEHPE 240  
QY 241 PKIHRMGDSFAVEIKGRIFHDFLPVVRRTINLPYITLAEVAVLGTGKSLGAEEIAAI 300  
Db 241 PKIHRMGDSFAVEIKGRIFHDFLPVVRRTINLPYITLAEVAVLGTGKSLGAEEIAAI 300  
QY 301 WETESMKLAQYMEDARATYELGKFFPMEASLAKLIQOSVWDVSRSTGNLVEYLL 360  
Db 301 WETESMKLAQYMEDARATYELGKFFPMEASLAKLIQOSVWDVSRSTGNLVEYLL 360  
QY 361 RVAVARNELAPNKDEEYRRLRTTYLGGVYKPEPERGLMENITYLDFRCLYPSIIVTHN 420  
Db 361 RVAVARNELAPNKDEEYRRLRTTYLGGVYKPEPERGLMENITYLDFRCLYPSIIVTHN 420  
QY 421 VSPDTLEBEGKNVDVPIVGYKCPDPPGIPISILGELITMROEIKKMKATIDPIEKK 480  
Db 421 VSPDTLEBEGKNVDVPIVGYKCPDPPGIPISILGELITMROEIKKMKATIDPIEKK 480  
QY 481 MLDVQRORAKLHA 493  
Db 481 MLDVQRORAKLHA 493  
QY 494 541 VNNLFAFSFNKKIKBSEVKVKALIRHKYKGAVEIQLSGKRKINITAGHSFTVRNGEI 600  
QY 494 601 KEVSGGIGEGDLIVAPKKIKLNEKGVSNIPILISLSEETADIVMTISAKGRKNFFK 660  
QY 494 721 SVKYNKREYLVMEINIKDFISYPPQKELEWKTGTNGFRTNCLKVDDEDFGKLLGY 780  
QY 494 781 VSEGYAGAKNKTGGISYVKLYNEDPNVLESKMKNVAEKKFKVVRVDRNCVSIKKNAYL 840  
QY 494 841 VMKCLCGALAEKNRIPSVILTSPEPVRSFLEAYTGGDIHPSKRFSLTSKELLANQL 900  
QY 494 901 VFLNLSLGISYVKIGFSDGVYRVYINEDLQFPQTSREKNTYNSNLI PKELLRDVFGKEFQ 960  
QY 494 961 KMTFKFKELVDYSGKLNREKAKLEFFINGDILVDRVKSVEKDEYGVYVDLSVEDNEN 1020  
QY 494 1021 FLVGFGLLYAHNSYGYMGYPKARWYSKCAESVTAWGRHYIEMTIRIEBKFGFKVLYA 542  
QY 543 D 1080

Db 1081 DSVGESEIIIRONGKIRFVKIKOLFQSKVDYSIGKEYCILEGVEALTLDGDKLVWKPV 1140  
QY 544 1141 PYVMEHRANKBMRILWLTNSWYIDVTDHSLIGYLTNTSKTKAKKIGERLKEVPFELGK 1200  
Db 544 1201 AVKSLICPNAPKDBNTKTSIAVFWELVGLVGNWGGDSRWAEYYLGLSTGKDAE 1260  
QY 544 1261 IKQKLEPLKTYGVISNYYYPKNEKGFNLAKSLVKFMKRFKDEKGRRKIPFMYELPV 1320  
Db 544 1321 TYIEAFLRGLFSADGTVTIRKGVPEIRLTNIDAFLEVRKLLWIVGISNISIFAETTPNR 1380  
QY 544 1381 YNGVSTGYYSKHLRIKKNRFAERIGFLIERKQELLEHLKSARVKRNTIDFGPDLVHVK 1440  
Db 544 1441 KVEEIPYEGYVYDIEVEETHRFFANNILVHNTDGFYATIPGKPELIKKAKEFLNYNS 1500  
QY 573 KLPGLLELEYEGFYLRGFFVAKRVAVIDEGRIITRGLVVRDWSIAKETQAKVLEA 632  
Db 1501 KLPGLLELEYEGFYLRGFFVAKRVAVIDEGRIITRGLVVRDWSIAKETQAKVLEA 1560  
QY 633 ILKESDVEKAVEIKVDVVEEIAKYQVPLEKLVHIEQITKQJSEYKAIQPHVAIAKRLAAK 692  
Db 1561 ILKESDVEKAVEIKVDVVEEIAKYQVPLEKLVHIEQITKQJSEYKAIQPHVAIAKRLAAK 1620  
QY 693 GKVPRTIISVILRGSKISDRVILLSEVDPKXHYDDPDYIENQVLPVLRILEAFG 752  
Db 1621 GKVPRTIISVILRGSKISDRVILLSEVDPKXHYDDPDYIENQVLPVLRILEAFG 1680  
QY 753 YRKEDLYOSSKQVGLDAWLKK 774  
Db 1681 YRKEDLYOSSKQVGLDAWLKK 1702

RESULT 2  
S67920  
DNA-directed DNA polymerase (EC 2.7.7.7) - Thermococcus sp.  
C;Species: Thermococcus sp.  
C;Date: 17-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 18-Jun-1999  
C;Accession: S67920  
R;Southworth, M.S.; Kong, H.; Kucera, R.B.; Ware, J.; Jannasch, H.W.; Perler, F.B.  
submitted to the EMBL Data Library, January 1996  
A;Description: Cloning, expression and modulation of the 3'-5' exonuclease activity of  
A;Reference number: S67920  
A;Accession: S67920  
A;Molecule type: DNA  
A;Residues: 1-775 <SOU>  
A;Cross-references: GB:U47108; NID:gl197451; PIDN:AAA88769.1; PID:gl197452  
A;Experimental source: strain 90N-7  
C;Superfamily: herpesvirus DNA-directed DNA polymerase  
C;Keywords: DNA binding, nucleotidyltransferase

Query Match 79.6%; Score 3204.5; DB 2; Length 775;  
Best Local Similarity 76.6%; Pred. No. 4.1e-157;  
Matches 592; Conservative 89; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIPDTYITKGGKPIIRIFKENGEPKIELDPHPQPIYALLKDDSAIEBIKAIGERHG 60  
Db 1 MILDVTIITENGKPIRVFKENGEPKIELDPTFTFYALLKDDSAIEDVKKVAKRGH 60  
QY 61 KIVRVDAVKKKFELGRDVEVWKLIFEHQDVPALRGKIREHPAVDIDYEYDIPAKY 120  
Db 61 TVVVKVGAEKVQKFLGRPIEVWKLIFENHPQDVPALRIRAHPAVDIYEYDIPAKY 120





Best Local Similarity 74.6%; Pred. No. 2,3e-154;  
Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

```
QY 1 MFDYDITKDGKPIIRIFKNGEFGKIELDPHFQPIYIALLKDDSAIDSIKAIKGRHG 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MILDADYITEGKPIIRIFKNGEFGKIELDPHFQPIYIALLKDDSAIDSIKAIKGRHG 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 KIVRVVDVAVKVKKFLGRDVEVWKLIFEPHPQDPVAPALRGKIREHPAVDIDYEYDIPFAKRY 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 KIVRVVDVAVKVKKFLGRDVEVWKLIFEPHPQDPVAPALRGKIREHPAVDIDYEYDIPFAKRY 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 LIDKGLIPMEGDEBELKMAFDIETFYHGDGDFGKGIIMISYADEEAEARVITKNIDILPY 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIDKGLIPMEGDEBELKMAFDIETFYHGDGDFGKGIIMISYADEEAEARVITKNIDILPY 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 VDVSNEREMIKFVQIVREKDPDLITVNGDNFDPYIKAEKLGVTLLGRDKEHPE 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VDVSNEREMIKFVQIVREKDPDLITVNGDNFDPYIKAEKLGVTLLGRDKEHPE 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 PKIHRMGDSFAVEIKGRTHFDLPFVVRTINLPYTLVAVYEAALGKSGVITKNIDILPY 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 PKIHRMGDSFAVEIKGRTHFDLPFVVRTINLPYTLVAVYEAALGKSGVITKNIDILPY 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 WETEESMKLAQYSMEDARATYELGKEFFPMEAEALAKIGOSVMDVSRSSTGNLVEWYLL 360
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 WETEESMKLAQYSMEDARATYELGKEFFPMEAEALAKIGOSVMDVSRSSTGNLVEWYLL 360
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 RVAYERNELAPNKPDEEYRRLRTYLGKYGKPERGLWENITVLDPRCLYPSIIVTHN 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 RVAYERNELAPNKPDEEYRRLRTYLGKYGKPERGLWENITVLDPRCLYPSIIVTHN 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 421 VSPDTLREGCKNVDVAPVIGVKFCDFPGFIPSLGELITWROEIKKOKATIDPIEKK 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 VSPDTLREGCKNVDVAPVIGVKFCDFPGFIPSLGELITWROEIKKOKATIDPIEKK 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 481 MLDYQORAVK----- 490
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 MLDYQORAVK----- 490
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 491 ---LHA----- 493
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 491 ---LHA----- 493
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 539 VAGIHAFSPDKSKKARVMAVKAVIRHRYSGNVIRVILNSGRKITTTEGHSFLVYRNGDL 598
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 VAGIHAFSPDKSKKARVMAVKAVIRHRYSGNVIRVILNSGRKITTTEGHSFLVYRNGDL 598
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 599 VEATGEDVXIGDILAVERSVNLPEKEERLNIIVELLNLNSPEETEDIILTIPIVKGKKNFPK 658
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 VEATGEDVXIGDILAVERSVNLPEKEERLNIIVELLNLNSPEETEDIILTIPIVKGKKNFPK 658
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 658 GMLRTLRWIFGEKRVFTASRYLRHLNLYIRLRKIGYDIDDKGEGKRYLYEKLVDV 718
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 658 GMLRTLRWIFGEKRVFTASRYLRHLNLYIRLRKIGYDIDDKGEGKRYLYEKLVDV 718
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

## RESULT 5

S68593  
DNA-directed DNA polymerase (EC 2.7.7.7) Vent, intein containing precursor - Pyrococcus  
N:Contains: DNA endonuclease (EC 3.1.1.-) PI-I; DNA-directed DNA polymerase (EC 2.7.7.7)  
C:Species: Pyrococcus sp.  
C:Date: 24-Aug-1996 #sequence\_revision 01-Nov-1996 #text\_change 24-Sep-1999  
C:Accession: S68593  
R:Xu, M.Q.; Southworth, M.W.; Marsha, F.B.; Hornstra, L.J.; Perler, F.B.  
submitted to the EMBL Data Library, August 1993  
A:Description: In vitro protein splicing of purified precursor and the identification of  
A:Reference number: S68593  
A:Accession: S68593  
A:Molecule type: DNA  
A:Residues: 1-1312 <XUA>  
A:Cross-references: EMBL:U00707; NID:G436492; PDB:AAA67130.1; PID:9825735  
R:Xu, M.Q.; Southworth, M.W.; Marsha, F.B.; Hornstra, L.J.; Perler, F.B.  
Cell 75, 1371-1377, 1993

```
QY 494 ----- 493
Db 719 VRNGNKRVLVEFNAVRDVISLMPBELKEMRIGTRNGFRMGTFVDIDEDFAKLLGYVY 778
QY 494 ----- 493
Db 779 SEGSAKWKQNTGGWSYTVRLYNENDEVLDLMEHLAKKFFGKVRGKNVVEIPKKNAYII 838
QY 494 ----- 493
Db 839 FESLGTLAENKRVPEVFTSSKGVWRWAFLEGYFIDGVDVHPKRVRLSTKSELLVNGIV 898
QY 494 ----- 493
Db 899 LLLNSLGSALILGSDGVYRVVNEELKFTFYRKKKNVYHSHVPKDILKFTFGKVPQK 958
QY 494 ----- 493
Db 959 NISYKKFRELIVNGKLDREKAKRIEVLNGDIVLDWRVBEIKREYYDYVYDLSVDEDNF 1018
QY 494 ----- 493
Db 1019 LAGFGLYAHNSYGYGYAKARWYCKEASVTAWGREYIEBFVRKELEKEGFKVLVID 1078
QY 544 TDCFYATIPGEXPETIKKAKFLKYNSKLPGLLELEYEGFYLGRGFFVAKRYAVIDEE 603
Db 1079 TDGLYATIPGAKPEEIKKALEFVDYINAKLPGLLELEYEGFYLGRGFFVTKKYYALIDEE 1138
QY 604 GRITTRGLEVRDMSIEIAKTOAKVLRAILKEDSVKAVEIVKDVVEIAKVQVPLEKL 663
Db 1139 GKIIITRGLVIRDRDSEIAKTOAKVLRAILKGVNEEAKVIVKEVTEKLSYIEIPPEKL 1198
QY 664 VIHQITKDSLEYKAGPHVAIAKRLAAGIKVRPTIISYIVLRGSGKISDRVILLSEY 723
Db 1199 VIYEQITRELPHEYKAGPHVAIAKRLAAGIKVRPTIISYIVLRGSGKISDRVILLSEY 1258
QY 724 DPKKKXKDPDYIENQVLPAVIRILEAFGRYKEDLYQSSKQVGLDNL 772
Db 1259 DLKHKYDAEYIENQVLPAVIRILEAFGRYKEDLYQSSKQVGLDNL 1307

RESULT 6
C71210
probable DNA-directed DNA polymerase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: C71210
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: C71210
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1235 <KAW>
A:Cross-references: GB:AP000007; NID:G3236134; PIDN:BAA31074.1; PID:G3258391
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1947
C:Superfamily: hypothetical protein PH0202

Query Match 71.7%; Score 2886; DB 2; Length 1235;
Best Local Similarity 46.5%; Pred. No. 1.7e-140;
Matches 574; Conservative 107; Mismatches 89; Indels 464; Gaps 5;

QY 1 MIFDIDYTKGKPIIRIFKKEGFKLEDPHFQPIYALLKDDSAIDKAIKGRHG 60
Db 1 MLDADYITDGGKPIIRIFKKEGFKLEDPHFQPIYALLKDDSAIDKAIKGRHG 60
QY 61 KIVRVVDVAVKKKFLGRDVEVWKLIFEPHPODVALRGKIREHPAVIDIYEYDIPFAKRY 120
```

```
Db 61 KVAIVITETKIQRFGLGRPIEVWKLYLEHPQDVPAIRDKIREHPAVVDIYEYDIPFAKRY 120
QY 121 LIDKGLIPMEGDEBELKMAPDIETFYHEGDFGKEIIMISYADEEARVITWKNIDLPY 180
Db 121 LIDKGLIPMEGNEKLTFLAVDITELYHEGDFGKEIIMISYADEEARVITWKNIDLPY 180
QY 181 VDVYSNEREMIKRQVQVIREKDPDLITYNGDNPDFLPYLKRAKPKLGVTLILLGRDKEHPE 240
Db 181 VEVVSSSREMIKRLIRVIKEDKPDVITTYNGDNPDFLPYLKRAKPKLGVTLILLGRD--NSE 238
QY 241 PKIHRMGDSFAVEITKGRHFDLPVVRRTINLPYTLLEAVYEAIVGLTKSKLGAEEATAAI 300
Db 239 PKQKMGDSLAVEIKGRHFDLPVVRRTINLPYTLLEAVYEAIVGLTKSKLGAEEATAAI 298
QY 301 WETESWKKLAQVSMEDARATYELGKEFFPWEAEALAKLICQSVMDVSRSTGNLVEWYLL 360
Db 299 WETGEGIERVAKYSMEDAKVTYELGREFFPWEAEALAKLICQSVMDVSRSTGNLVEWYLL 358
QY 361 RVAYERNELAPNKPDEEYERRLRTTYLGGYKPEPERGLWENITYLDLFRCLYPSIIVTHN 420
Db 359 RYAYERNELAPNKPDEEYERRLRTTYLGGYKPEPERGLWENITYLDLFRCLYPSIIVTHN 418
QY 421 VSPDTLREGCKNYDVAPIVGYKFCDFPGFIPSIILGELITMRQEIKKMKATIDPIEKK 480
Db 419 VSPDTLNREGCEYDVAPKVGHRFCDFPGFIPSIILGELITMRQEIKKMKATIDPIEKK 478
QY 481 MLDYRQRAVKL----- 491
Db 479 LLDYRQRAIKILANSILPDEWLPFIVENKRVFKIGDFIDREIEENAEVRKRDGETEILE 538
QY 492 ----- 491
Db 539 VKDLKALSPNRETCKSELKKVXKALIRHRYSGKYSIKLSGRBIKITSGHSLPSVXNGKL 598
QY 492 ----- 491
Db 599 VKVRGDELKPGDLVVVPGRLKLPESQVNLVELLLKLPESETSNTIWMIPVAKRKNFFK 658
QY 492 ----- 491
Db 659 GMLKTLWIYFEGGERPRTAGRYKHLERLGYVKLKERGCCEVLDWESLKRKYKLYETLKN 718
QY 492 ----- 491
Db 719 LKYNNGSRAYMVFBNLSRDVSLMPTEELKEMIIGEPRGPKIGTFIDVDDSPAKLLGYVI 778
QY 492 ----- 493
Db 779 SSGDVEKDRVKFHSKQNVLEDAKLAELFKVRGRGRIEYVSGKISHAIFRVLAEGRK 838
QY 494 ----- 493
Db 839 IPEFIFTSPMDIKVAPLKGNGNABELTFTSKSELLVNLILLNSIGVSDIKIEHKGK 898
QY 494 ----- 493
Db 899 YRVYINKSSNGDIVLDSVESIEVEKVEGYVDVSVEDNENFLVGFGLLYAHNSYGYV 958
QY 501 GYPKARWYKCAESVTAWGRHYIEMTKEIEKEFGKVLVYADTDGFYATIPGEXKPE-ETI 559
Db 959 GYAKARWYKCAESVTAWGRQYIDVRRLEAR-GFKVLYITDGLYATIPGEXKPEV 1017
QY 560 KKKAKEFLKYNKSLPGLLELEYEGFYLGRGFFVAKRYAVIDEGRITRGLVVRDWS 619
Db 1018 KRRALFEVDYINSLPGLLELEYEGFYLGRGFFVAKRYAVIDEGRITRGLVVRDWS 1077
QY 620 EIAKTEQAKVLBAIKEDSVKAVIRVDVVEIAKVQVPLEKVIHQITKDSLEYKAI 679
Db 1078 EIAKTEQARVLEAILKHGNEEAVKIVKDVTEKLTNYEVPPEKLVYEQITRFPINEYKAI 1137
QY 680 GPHVAIAKRLAAGIKVRPTIISYIVLRGSGKISDRVILLSEYDIPKHKYDIPDYINQ 739
```

Db 1138 GHVAVAKLMARGIKVKGWIGVILRGDGPISKRAISIEEFPDRKHKYDAEYIENQ 1197

QY 740 VLPVAVRILAEQYRKEDLYQSSKQVGLDAWLK 773

Db 1198 VLPVAVRILAEQYRKEDLYQSSKQVGLDAWLK 1231

RESULT 7

S71551

DNA-directed DNA polymerase (EC 2.7.7.7) KOD, intein containing precursor - Pyrococcus s

N;Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA endonuclease (EC 3.1.-.-) PI-II; DNA

C;Species: Pyrococcus sp.

A;Variety: strain KOD1

C;Date: 22-Nov-1996 #sequence\_revision 22-Nov-1996 #text\_change 20-Apr-2000

C;Accession: S71551

R;Kakihara, H.; Takagi, M.; Imanaka, T.

A;Description: Cloning and analysis of the DNA polymerase gene from a new hyperthermophi

A;Reference number: S71551

A;Accession: S71551

A;Molecule type: DNA

A;Residues: 1-1670 <KAK>

A;Cross-references: EMBL:D29671

C;Function: <NUCL>

A;Note: DNA-directed DNA polymerase, catalyzes the polymerization of DNA at th

A;Function: <EN1>

A;Description: as DNA endonuclease PI-I, catalyzes the hydrolysis of internal phosphodie

A;Note: DNA endonuclease PI-PspI

C;Function: <EN2>

A;Description: as DNA endonuclease PI-II, catalyzes the hydrolysis of internal phosphodi

A;Note: DNA endonuclease PI-PspII

C;Superfamily: DNA-directed DNA polymerase KOD

C;Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing

F:1-406/767-851/1388-1670/Product: DNA-directed DNA polymerase KOD #status predicted <NA

F:1-406/Domain: DNA-directed DNA polymerase KOD extein 1 #status predicted <XT1>

F:407-766/Product: DNA endonuclease PI-I (pol KOD extein 1) #status predicted <MAT2>

F:767-851/Domain: DNA-directed DNA polymerase KOD extein 2 #status predicted <XT2>

F:852-1387/Product: DNA endonuclease PI-II (pol KOD extein 2) #status predicted <MAT3>

F:1388-1670/Domain: DNA-directed DNA polymerase KOD extein 3 #status predicted <XT3>

F:406-767/Cross-link: peptide (Arg-Ser) #status predicted

F:851-1388/Cross-link: peptide (Asn-Ser) #status predicted

Query Match 68.7%; Score 2764.5; DB 2; Length 1670;

Best Local Similarity 35.9%; Pred. No. 4.5e-134;

Matches 599; Conservative 83; Mismatches 88; Indels 899; Gaps 4;

QY 1 MIFDTYITKDGKPIIRIFKKEGFEKIEIDPHFPQPIYIALLKDDSAIDBIKAIGERHG 60

Db 1 MILDTYITKDGKPIIRIFKKEGFEKIEIDPHFPQPIYIALLKDDSAIDBIKAIGERHG 60

QY 61 KIVRVDAVKKKFLGRDVEVWKLIFEPQDVPAIRGKIREHPAVIDIYEYDIPFAKRY 120

Db 61 TVVTYKRVKQKFLGRDVEVWKLIFEPQDVPAIRGKIREHPAVIDIYEYDIPFAKRY 120

QY 121 LIDKGLIPWEGDEELKMAFQIETFYHGBDFGKEGIMISYADEEAEARVITWKNIDLPY 180

Db 121 LIDKGLIPWEGDEELKMAFQIETFYHGBDFGKEGIMISYADEEAEARVITWKNIDLPY 180

QY 181 VDVVSNREMIKRFQIVREKDPDVLITYNGNDPLPYLIRKAGSKLGVTLILGRDKHPE 240

Db 181 VDVVSTEREMIKRFLRVVKEKDPDVLITYNGNDPLPYLIRKAGSKLGVTLILGRDKHPE 240

QY 241 PKIHRMGDSFAVEIKGRIFHFDLPVVRRTINLPTVLTAVYEAVALGKTKSLGAEIAAI 300

Db 239 PKIQGRDFRFAVEKGRIFHFDLPVVRRTINLPTVLTAVYEAVALGKTKSLGAEIAAI 298

QY 301 WTETSMKKLAQYSMEDARATYELGKEFFPMELAKLIGOSVMDVSRSSSTGNLVEWYLL 360

Db 299 WETGENLVARYSMEDAKVYELGKEFFPMELAKLIGOSVMDVSRSSSTGNLVEWYLL 358

QY 361 RVAYERNELAPNKDEKELARR-RQSYEGYVKEPERGLWENIVYLDLFRCHPADTKVWVK 410

Db 359 RKAYERNELAPNKDEKELARR-RQSYEGYVKEPERGLWENIVYLDLFRCHPADTKVWVK 417

QY 411 ----- 410

Db 418 GKGIINISEVQEGDYVLGIDGQWRVQRWYDYKGBELVNINGLCKTPNHHKLPVVTIKNERQ 477

QY 411 ----- 410

Db 478 TRIEDSLAKSFLTKVKGKIITPFLVEIGRATSENIPEEVLKGBELAGILLAEGLLRK 537

QY 411 ----- 410

Db 538 DVEYFDSSRRKGRIRSHQYRVEITIGKDEEFRDRITYIFERLFGITPFSSEKKGTVAVTL 597

QY 411 ----- 410

Db 598 KVAKKNVYLKVEIMDNIESLHAPSVLRGFFEGDGSVNRVRSIVATQGTQKNEWKIKVS 657

QY 411 ----- 410

Db 658 KLSQLGIPHQTYTYQYQENGKDRSYILEITGKGLILFQTLIGFISEKKNALLKKAIS 717

QY 411 ----- 420

Db 718 QRENNLENGFYLSFNFVSTVEYEGKVYDVLTLGTPYFANGILTHNSLYPSIITHN 777

QY 421 VSPDTEREGCKNYDVAPIVGYKCPDPPGIFPSIIGELITMRQBIKKQKATIDPIEKX 480

Db 778 VSPDTLANREGCKEYDVAPQVGHRECKDPGPFIPSLGLLEERQKIKKQKATIDPIERK 837

QY 481 MLDYRQRAVKLHA----- 493

Db 838 LLDYRQRAIKITANSILPEHMLPVLEGEVHFVRIIGELIDRMMEENAGKVKEGETEVL 897

QY 494 ----- 493

Db 898 VSGLEVPSPNRTNKALRKVKALIRHDYSGKVTIRLKSRRIKITSCHSLPSVRNGEL 957

QY 494 ----- 493

Db 958 VEVTGDELKPGDLVAVPRRLBELPERNVHLNVELLGTPEETLDIVMTIPVKCKQNF 1017

QY 494 ----- 493

Db 1018 GMLRLRWIFGEEKRPTARRYLRLHLEDLYVRLKIGYEVLDWDSLNKRYRLYEALVEN 1077

QY 494 ----- 493

Db 1078 VRYNGNKEEYLVEFNSIRDAVGIMPLKELKEWKIGTLNGFRMRKLI EVDLSAKLLGYV 1137

QY 494 ----- 493

Db 1138 SEGYARKQNPKNQWSYVKLYNEDEPVLDDMERLASRFFGKVRGRNRYVEIPKKIGYLL 1197

QY 494 ----- 493

Db 1198 FENMCGVLAENKRIFEFVTPSPKGVRLAFLEGYSAMATSETEQATQALNEKRALANQLVL 1257

QY 494 ----- 493

Db 1258 LLNSVGVSAVKLGHDSGVYRVINEELPFVKLDKKKNAYSHVPIKEVLSEVFGVKFQKN 1317

QY 494 ----- 493

Db 1318 VSPQTRKMWEDGRLDPEKAQRLSWLIEGDVVYLDREVSDYDGYVYDLSVENENFL 1377

QY 494 ----- 544

Db 1378 VFGGLVANKNSYGYGVARARWYCKEASVTANGREVIITWIKIEIKYGFVYSDT 1437

QY 545 DGFYATIPGKEPETTKKAKFLKYNISKLPGLLEIEYEGFVLRGFFVAKRYAVIDSEG 604

Db 1438 DGFATIPGADAETVKKAMEFLNINAKLPGALEIEYEGFVLRGFFVTKKRYAVIDSEG 1497

QY 605 RITTRGLEVVRDMSIAKETAQVLEAILKSDSVEKAVEIKOVVBEIAKYQVPLEKLV 664  
 Db 1498 KITTTRGLETVRRDMSIAKETAQVLEAILKSDSVEKAVEIKOVVBEIAKYQVPLEKLV 1557  
 QY 665 IHEQITKDLSEYKATGPHVAIAKLAAGIKVRPGTIIISYIVLRGSGKISDRVILLSEYD 724  
 Db 1558 IHEQITRDLKDYKATGPHVAVAKLAARGVKIRPGTIVISYIVLRGSGKISDRVILLSEYD 1617  
 QY 725 PKGHKYDPDYIENQVLPVAVLILAFAGYRKEDLKYQSKQVGLDAMLK 773  
 Db 1618 PTKHKYDAEYIENQVLPVAVLILAFAGYRKEDLKYQSKQVGLDAMLK 1666

RESULT 8  
 A69312  
 DNA polymerase B1 (polB) homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: A69312  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: A69312  
 A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-781 <KLE>  
 A:Cross-references: GB:AE001070; GB:AE000782; NID:G2689393; PIDN:AAB90741.1; PID:G265013  
 C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 32.3%; Score 1301; DB 2; Length 781;  
 Best Local Similarity 37.4%; Pred. No. 2.3e-59;  
 Matches 305; Conservative 155; Mismatches 266; Indels 90; Gaps 21;

QY 2 IFDTYITKDGPIIRIFKKEGFEKIELDPHFQPIYIALLKDDSAIDEIKRGERGK 61  
 Db 8 LIDADYETIGKAVVRLWCKDQGIQFVAYDNFDYFVIGVDE---DILKNAATSRRE 64  
 QY 62 IYRVVDVAVKKKFGVGRDVEWVKLIFEHPDVPALRGKIREHPAVIDIYEDIPAKRYL 121  
 Db 65 VIKLKSFEKAQKLTGREVEGVIVYAHHPQHVFKLRYLSQFG---DVREADIPPAKRYL 121  
 QY 122 IDKGL-----IPWEGD-----ELKLMAPDIETFYHEG-DBFG 153  
 Db 122 IDKDLACMDGIAIEGEGKGGVIRSYKIEKVEIPRMEFPELKMVLFDCEMLSSFGMEPE 181  
 QY 154 KGEIIMISVADDEEARVITWKNIDLPYDVVSNREMIKRFQIVIREKDPVLTLYNGDN 213  
 Db 182 KDPILVISVKTNDDEII-----LTGDERKIISDFVKLTSDYDPIIVGYNQDA 230  
 QY 214 FDLPLYIKRAEKLGVTLILGRDKEPPEPKIRHMGDSFAVEIKGRIFHDFLPVVERTINLP 273  
 Db 231 FDMPLYLRKEAERWNTPLDVGRGDSN---VVFEGG---RPKITGRLVNDVLIAMRISDIK 284  
 QY 274 TTYLAEVAVLGKTK-SKLGAEEIAAWEETESKKLAQVSMEDARATYELGKEFFPME 332  
 Db 285 IKKLENAVEFLTKIEADIEAKDIYRYWSRGEK-EKLVYARQDAINTYLIKELLPMH 343  
 QY 333 AELAKLIGOSVMDVSRSTGNLVENVLLRVAVERNELAPNKPDEEYRRLRTYLGQVY 392  
 Db 344 YELSKWIRLPVDVTRMGKQVDMILLSEAKKIGEIAPNPEHAE-----SYEGAFV 396  
 QY 393 KPEPGLMNTIYLPFRCLYPSIIVTHVNSPDTLRECKN--YDVAPIVGYKFKCKPPG 450  
 Db 397 LEFERGLHENVACLDFAFMYPSIMIAFNISPTY---GCRDDCYE-APEVGHKFKRSPDG 452  
 QY 451 FIPSIILGELITRQRIKKWK-ATDPLEKMLDRQRAVKLHANSYGYGYGPKARWYS 509

Db 453 FFKRILRLMLIEKRELKVELKNLSPSSSEYKLLDIKQOTLKVLNTNFGYGMGNLARWYC 512  
 QY 510 KECAESVTAGRVHVIEMTIKIEIEKPGKVLVADTDFGFIATIPGEKPEITIKKAKSEFLKY 569  
 Db 513 HPCAEATTANGRFITSAK-IAESMGFKVLYGDTDSIFVTKAGMTYKEDVDR-----LIDK 567  
 QY 570 INSKLPGLLEVEYGFYLRGFFVAKRYAVIDBEGITRTRGLEVVRDRDSEIAKETQAKY 629  
 Db 568 LHBEPL--IQIEVDEYSAIFFVEKRYAGLTBDGRVLVWKGLEVRVRGDMCELAKYQREV 625  
 QY 630 LEAILKEDSVEKAVEIKOVVBEIAKYQVPLEKLVHEQITKDLSEYKATGPHVATAKRL 669  
 Db 626 IEVILKEKPEKALSIVKQVILRIKESKVSLEEVYIKGLTKPKSKYESQAHVKAALKA 685  
 QY 690 AAKGKIVRPGTIIISYIVLRGSGKISDR---VILLSEYDPKK-----HKYDPDY 735  
 Db 686 REMGIIYPVSSKIGYIVKSGNIGDRAPVLDLIEDFDGLENLRITKSGTEIKKLDKDY 745  
 QY 736 IENQVLPVAVLILAFAGYRKEDLKYQSKQVGLDAMLK 771  
 Db 746 IDNQIIPVLRILRERFGYTEASLK--GSSQMSLDSF 779

RESULT 9  
 E64410  
 DNA-directed DNA polymerase (EC 2.7.7.7) family B, intein containing precursor - Methan  
 N:Contains: DNA endonuclease (EC 3.1.1.1) PI-I; DNA endonuclease (EC 3.1.1.1) PI-II; DN  
 C:Species: Methanococcus jannaschii  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 13-Sep-1998  
 C:Accession: E64410  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasch  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: E64410  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1634 <RUL>  
 A:Cross-references: GB:U67532; GB:L77117; NID:G1591559; PID:G1591563; TIGR:MJ0885; PID:  
 C:Genetics:  
 A:Map position: REV816304-811400  
 A:Start codon: TTG  
 C:Function: <DPL>  
 A:Description: as DNA-directed DNA polymerase, catalyzes the polymerization of DNA at t  
 C:Function: <EN1>  
 A:Description: as DNA endonuclease PI-MjAI, catalyzes the hydrolysis of internal phosph  
 C:Function: <EN2>  
 A:Description: as DNA endonuclease PI-MjAI, catalyzes the hydrolysis of internal phosph  
 C:Superfamily: DNA-directed DNA polymerase KOD  
 C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing  
 F:1-425,795-882,1359-1634/Product: DNA-directed DNA polymerase family B extein 1 #status predicted <XT1>  
 F:1-425/DNA: DNA-directed DNA polymerase PI-I (pol B extein 1) #status predicted <XT1>  
 F:426-794/Product: DNA endonuclease PI-I (pol B extein 1) #status predicted <XT2>  
 F:795-882/DNA: DNA-directed DNA polymerase family B extein 2 #status predicted <XT2>  
 F:883-1358/Product: DNA endonuclease PI-II (pol B extein 2) #status predicted <XT2>  
 F:1358-1634/DNA: DNA-directed DNA polymerase family B extein 3 #status predicted <XT3>  
 F:425-795/Cross-link: peptide (Arg-Ser) #status predicted  
 F:882-1359/Cross-link: peptide (Asn-Ser) #status predicted

Query Match 32.1%; Score 1291; DB 2; Length 1634;  
 Best Local Similarity 22.7%; Pred. No. 2e-58;  
 Matches 374; Conservative 159; Mismatches 217; Indels 894; Gaps 23;

QY 2 IFDTYITKDGPIIRIFKKEGFEKIELDPHFQPIYIALLKDDSAIDEIK-AIK 55  
 Db 13 LIDNTYKTTEDKAVIYVLINS---ILKDRDFKPYFVELHKEKVENEDIEKKEFLK 68  
 QY 56 GERHGKIVRVDAVKYKK-FLGRDVEWVKLIFEHPDVPALRGKIREHPAVIDIYED 114  
 Db 69 ND-----LLKPVENIEVVKIILRKEKEVIKIATHPQKPKLR-KIKCEIYKEIYEHDI 123

115 PFAKRYLIDKGLIPM-----EGDE-----ELKMAFDIETFYHEGDFGKE-IIMIS 161  
 124 PFAKRYLIDNEIIPMTYWFENKQVSIPIPKLSVAFDMEVYNRDTEPNPDRPILMAS 183  
 162 YADEEARVITTKNIDLPVYVVSNEREMIKRQVIREKDPDVLITTYNGDNFDLPYLK 221  
 184 FWDENGKVITYEFNHPNIEVUVNKEKLIKLIETLKEY--DVIITYNGDNFDLPYLKA 241  
 222 RAELKGVTLILGRDKEHPEKIHMRGDSFAVEIKGRIHFDLPVVRTINLPTYLEAVY 281  
 242 RAKTYGIDINLKDGE--ELKIRGGMEVRSYIFGRVHDLVPISRRLKLTAKYLEDVV 299  
 282 EAVLGKTKSLGABEIAAIWETESMKLAQYSMEDARATYELGKEFFPMZAEALAKIQ 341  
 300 YNIFGIEKLIPTKIVDYVAND--KTLIEYSLQAKYTYKIGKYFFPLEVMSFIVNQ 357  
 342 SVMDSRSSTGNLVEWYLLVAVARNELAPNKDPEBEYRRRLTTLTGGYVKEPPEGLWE 401  
 358 TPPEITRMSSQMVYLLMKRAFKNMIVENKDEBEYRRRLTTLTGGYVKEPPEGLWE 417  
 402 NITVLDPRC----- 410  
 418 DILSMDFRCHPKGTGVVVKGVINIEDVKEGNYVLGDQWQKVKWVYEYEGELINVN 477  
 411 ----- 410  
 478 GLKCTPNHKIPLRYKIKHKKINKNDYLVRIYAKSLTLTKFGEGLILCKDPETIGNYEK 537  
 411 ----- 410  
 538 YINDMDEDFILKSELIGILAEGLLRDIEYFDSRGKKRISHQYRVEITVNEDEKDFI 597  
 411 ----- 410  
 598 EKIRYIFKLPNYELYVRRKGTAKITLGCACKDIYLIKBEILKNKEKLPNAILRGFFE 657  
 411 ----- 410  
 658 GDGYNTVTRAVVNOGNTNVDKIFIASLLDRIGIKYSFTYSYBERGKLRKYVIEIF 717  
 411 ----- 410  
 718 SKGDLIKPSILISFISRRKNLLNEIIRKTLYKIGDYGVYDLDDVCVSLESYKGEVYDL 777  
 411 -----LYPSIIVTHNVSPDTEREGCKNYDA-PIVGKFKCKPFGF 451  
 778 TLEGPPYFANGILTHNSLYPSIISYNISPTLDCCECK--DVSEKILGHWFCKKEGL 835  
 452 IPSILGELITWRQEKKKMK--ATIDPI--EKKMLDYRQRAVKLHA----- 493  
 836 IPKTLRNLIERINIKRKKWMAEIGEINEEVNLLDYEQSKLILANSILPDEYLTITEE 895  
 494 ----- 493  
 896 DGIKVVKIGEYIDDLMRKHDKIKFSGISEILETKNLKTFSDFKITKKEIKKVKALIRH 955  
 494 ----- 493  
 956 PYFGKAYKILRSRTIKVTRGSHLFPKYENGKIVEKGDDYRFGDLIVVPKKLTCVDREV 1015  
 494 ----- 493  
 1016 VINIPKRLINADEEIKDLVITKHKDKAFFVKLKTLEDIENKUKVIFDDCILYKELG 1075  
 494 ----- 493  
 1076 LIDYNTIINKVNDIKILDEEKFAKKYFDVIBHGNFKKRCNIQIKIKDYIANIPD 1135  
 494 ----- 493  
 1136 KEPEDCEIGAYSGINALLKDEKLAKFLGFFVTRGRLLKQKLGKETVYISVYKSLPEY 1195

494 ----- 493  
 1196 QKEIAETFEVFGAGSMVKDVTMDNKIVVLKYIFKCGDKDKKHIPBELFLASBSVIK 1255  
 494 ----- 493  
 1256 SFLDGLKAKNSHKGTSTFMAKDEKYNOLMILFNLVGIPTRTPVONKGYKLTLPKY 1315  
 494 -----NSYGYMGYPKARWYSKE 511  
 1316 GTVKDLMDLVEKEIEAFEYSYGVYVYDLSVEDNENFLVNNIVAHNSVGYGLAFPRARFYSRE 1375  
 512 CAESVTAWGRHYIMTKEIEEKGFKVLYADTDGFIATIPGEK--PETIKKAKPELKY 569  
 1376 CABITVTLGRKYLITETVKE-AEKGFKVLYDIDGFIYA--IWKEKISKEELIKKAMEFVY 1433  
 570 INSKLPGLLELEYEGFYLRGFFVAKRYAVIDEGRITTRGLEVVRDSEIAKETQAKY 629  
 1434 INSKLPGLMELEFEGYFKRGIFVTKRYALIDENGRVTKGLEFVRDWSNIAKITQRRV 1493  
 630 LEALIKESVKAVEIKVDVVEEIAKQVPLEKVIHQETKDLSEYKATGPHVAIAKRL 689  
 1494 LEALLVEGSIKAKKIIQDVIKDLREKKIKEDLIIYTQJTKDPKRYTAPHVEIAAKL 1553  
 690 AAKGKIVRPGTIIISVILRGSGKISDRVILSEYDPKKHKYDPPDYIENQVLPVLRILE 749  
 1554 MREGKIKVGDIIIGYIVKGTKSISERAKLPEEVD--IDDIDVNYIDNQILPPVLRIME 1611  
 750 AFGYKEDKLYOSSKQVGLDAWLK 773  
 1612 AVGSKNELKEGA-QLTLDKFFK 1634

RESULT 10  
 C69028  
 DNA-dependent DNA polymerase family B (PolB1) - Methanobacterium thermoautotrophicum (C) Species: Methanobacterium thermoautotrophicum  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C/Accession: C69028  
 R/Smith, D.B.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qi, S.; Church, G.M.; Daniele, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fun  
 A/Reference number: A69000; MUID:98037514; PMID:9371463  
 A/Accession: C69028  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-586 <MTH>  
 A/Cross-references: GB:AE000888; GB:AE000666; NID:G2622304; PIDN:AA885697.1; PID:G2622  
 A/Experimental source: strain Delta H  
 C/Genetics:  
 A/Gene: MTH1208

Query Match 30.0%; Score 1209.5; DB 2; Length 586;  
 Best Local Similarity 42.9%; Pred. No. 7.7e-55;  
 Matches 255; Conservative 107; Mismatches 176; Indels 57; Gaps 10;

1 MIFDPTDITKDGKPIIRIFKE---NGEFKIELDPHFQPIYIALLKD-DSAIDEIKAKG 56  
 7 VLLDIDVTVVDEVPVIRLFQKDKSGNEPIIAHDSFRFYIYAIPTDLDCELELEEL 66  
 57 ERHGKIVRVVDVAVKVK-KFLGRDVEVWKLIFTEHPQDVEALRGKIREHPAVIDIYEYDIP 115  
 67 EK-----LEVEMEDLGRPTVIRIEFPHQDVPKIRDIRDLESVRDIREHDIP 116  
 116 FAKRYLIDKGLIPME-----GDEELKLMARFI 142  
 117 FYRRYLIDKISIVEMEELFQGEVDSAPSVTTDRTVETVGRVQSGGAHGLDILSFDI 176  
 143 ETYFHEG-DFGKGEIIMISYA-DEBEARVITKNIDLPVYVVSNEREMIKRQVIRE 200  
 177 EVRNPHGMPPDEKDEIVMIGVAGNMGVESVISTAGDHLDFVEVVEDERELLEFAFVID 236



130 -----EGD-----BELKMAPDIETFYH-EGDEF 152  
124 MYAEVREAVPHGYSDRAYTLSDIREDETRIOEDPLKGLRVMAPDIEVYSKORPDP 183  
153 KGGEIIMISYAD-----EEEARVITWKNIDLPYVDVWNSNEREMIKRFYQIVREKOPD 204  
184 KDPVIMIGLOQAGGIEILEAEDR-----SDKKVIAGFVERVKSIDPD 227  
205 VLIITNGONFDLPYILKRAEKLGVTLGLGRKEHPEPKLHMGDSFAVEIKGRHIFDLFP 264  
228 VTIVGYNQRFDFWYUVERARVLGVKAVGRSVEPQGLYG-----HYVSGRLNVLD 282  
265 VVRRTINLPTYLEAVEAVLGKTK--SKLGAE--EIAAIWETEEBSMKLAQSYMEDARA 320  
283 FAEELHEVVKYLEEAD-YLGVKIGERVLTLEWQICEYWDPSKRLAKYLRDDVKS 341  
321 TYELAKPEFPMBAELAKLIGOSVWDVSSSTGNLVEVTLRLVAVERNELAKPKDEEYR 380  
342 TMGLAEKFLPFCALSSQVGLPDCQMAASVGFLEWRLIREAAKLGSLVNVNRSRGR 401  
381 RLRLTYLGGYKKEPERGLWENITVDFCLYPSIIVTHNVSPDTLREGCKNYD----- 435  
402 -----YAGAVLRPKGVHEDIAVLDFASVPMVKNVNGVDPDLVRPG-EEYGEVEY 454  
436 VAPIVGYKCDPPGFIPISILGELITMRQEIKKOKKA-TIDIEKMDLYQRAVKLHAN 494  
455 TAPVGHKPRKPPGFFKILERFLSRQIRSEMKKPPDPSPYKLLDERQKAIKLAN 514  
495 SVYGYGYPKAWYKCAESVTANGRHVIENTIKEIEKFGKVLXADTDCGYATIPGE 554  
515 ASVYGNWPHAWYCRECAEAVTANGRSITIRAIKAGE-LGLEVIYGDTSLSFVKNDPE 573  
555 KPETIKKAKFEIKYINSKLPGLLEVEGFGYLRGFFV-AKRYAVIDEEGRIITRGLEV 613  
574 KVERLIRFVEELGF-----DIKVDKVRVRFVFEAKRYVGLTVDGRKIDVVGFBA 624  
614 VRDWSIAKETQAKVLEAILKEDSVKAVEIVKDVESIAKYQVPLEKVIHQITKDL 673  
625 VRDWSIAKETQAKVLEAILKEDSVKAVEIVKDVESIAKYQVPLEKVIHQITKDL 684  
674 SEYKAIQPHVAIAKLAAGIKVRPOTIISYIVLRGSKISDRVILLSYDPKHKYKDP 733  
685 SMYEAQPHVTAALLMERAGIKVEPGAKIGYVTKSGPLYTRA--KPYFMASKEEVDVE 742  
734 YYTENQVLPVAILAEAFGRKEDLYKQSSKQVGLD 769  
743 YYVDKQVPAALRILOFYGVTEKRLKGGGRQSTLLD 778  
RESULT 14  
T05731  
DNA-directed DNA polymerase (EC 2.7.7.7) delta chain - soybean  
C:Species: Glycine max (soybean)  
C>Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 18-Jun-1999  
C/Accession: T05731  
R:Collins, J.T.B.; Cannon, G.C.; Heinrichst, S.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: Z15439  
A:Accession: T05731  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1088 <COL>  
A:Cross-references: EMBL:AF020193; NID:G2895197; PIDN:AAC18443.1; PID:G2895198  
C:Genetics:  
C:Gene: Pol delta  
C:Function:  
A:Description: catalyzes replication of DNA  
C:Superfamily: herpesvirus DNA-directed DNA polymerase  
C:Keywords: DNA binding; DNA replication; nucleotidyltransferase  
Query Match 17.2%; Score 693.5; DB 2; Length 1088;  
Best Local Similarity 27.2%; Pred. No. 5.5e-28;  
Matches 237; Conservative 149; Mismatches 324; Indels 161; Gaps 30;

226 -DRVREFVYRAAFDPIITVYNSHDFWYLMERARRLGKLDVTR-RVGAETTSV 283  
246 MGSFAYEIKGRIHFDLPVVRRTINLPTVLEAVEAVLGKTKSLGAE--IAAIWE 302  
284 YGH---VSVQRLNLDYDAEMPELIMKTLLEAVEAVLGKTKSERVIEWRIFBYWD 340  
303 TEEMKLAQYSMEDARATYELGKEFPFMEAEAKLIGOSVWDVSSSTGNLVEVTLRLV 362  
341 DEKKQLLERYALDVRATYGLAEKMLPFAQLSTVTVGPDLQVGMGVGFRLIEWYLMRA 400  
363 AYERNELAPKPEDEEYRRRLRTTYLGGYKKEPERGLWENITVDFCLYPSIIVTHNV 422  
401 AYDMNELPNRVE-----RGEYSKGAIVLPKGVHENVVVDJFSMTSPSIMIKYNG 454  
423 PDTL--ERECKNYD---VAPIVGYKCDPPGFIPISILGELITMRQEIKKOKKA-TIDP 476  
455 PDTIVDDSECPKYGGCVAVEVGHRRPSPGFFKTVLENLLKRLRQVKEKKEFPDPDS 514  
477 LEKMDLYQRAVKLHANSYGYWYKAWYKCAESVTANGRHVIENTIKEIEBKFG 536  
515 PEYRLYDEROKALKVLANSYGYWYKAWYKCAESVTANGRHVIENTIKEIEBKFG 573  
537 FKVLYADTDGFIATIPGKPEIKKAKFEIKYINSKLPGLLEVEGFGYLRGFFV-AKK 595  
574 LKVIYGDTSLSFVVD-----KEKVEKLEFVEKELG--FEIKIDKIYKKVFFTEAKK 624  
596 RYAVIDEGRITRGLVVRDWSIAKETQAKVLEAILKEDSVKAVEIVKDVESIAK 655  
625 RYVGLLEDGRIDIVGFAVRGDMCELAKEVQEAABIVLNTGNVDKALSYIREVIKQRE 684  
656 YQVPLEKVIHQITKOLSEYKAIQPHVAIAKLAAGIKVRPOTIISYIVLRGSKISD 715  
685 GKVPITKLIIMKTLKRIEVEHDAPHVWAARMKEAGYVSPGDKGVYIVKSGSVSS 744  
716 RVLLSEVDPKHKYKDPDYIENQVLPVAILAEAFGRKEDLYKQSSKQVGL 768  
745 RAYPFVMDPS--TIDVNYIDHQIVPAALRIISYFVTEKQLKAAATVQVRS 795  
RESULT 13  
E72515  
probable DNA-directed DNA polymerase APE2098 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C/Accession: E72515  
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: E72515  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-784 <XAW>  
A:Cross-references: DDBJ:AP000063; NID:G5105654; PIDN:BAA81109.1; PID:G5105797  
A:Experimental source: strain K1  
C:Genetics:  
C:Gene: APE2098  
C:Superfamily: herpesvirus DNA-directed DNA polymerase  
Query Match 28.3%; Score 1138; DB 2; Length 784;  
Best Local Similarity 34.6%; Pred. No. 5.3e-51;  
Matches 282; Conservative 137; Mismatches 287; Indels 110; Gaps 19;  
11 DGKPIIRIFKNGEKFKELDPHFQYIYALLKDDSAIDEIKAKGERHGKIVRVVDAVK 70  
16 DGSRVVVFY----GEFR-----PFYFVLPDGSVGLDQLAAMIRLRSRSPSPILSVR 63  
71 VKKKEIGRDVYKWLIFEPHODVPALRGKIREHPAVIDIVEYDIPPAKRYLIDKGLIPM- 129  
64 VRRRFIGREVALKVTTLVPASVREYREAVRLLGGVDRVLEADIPPALRFIIDFNLYPMR 123



Qy	15	IIIRFKENGEGF	KIELDPH-POPIYIYALLKODS	SAIDEI-----KAJGERHGK	61
Dd	88	IIIIFGVYTBEGH	SVCCNVGFBFYIICCPOMG	PDDISHPHOTLEGREAEVRNS	147
Qy	62	IVRVDAVKYK-----	KKFLGRDVEVWKLIFEHPQ	VPALRGKIEHPAVID-----	108
Dd	148	FVRIEMVQRRS	IMYYQNSQPF-----	KIVVALPTWVASCGR-ILDRGI	199
Qy	109	-----IYVEDIP	FAKXYLIDKGLI-----		127
Dd	200	MKSFLYIENVL	FAURFMDICNVGNWIGIP	AGKYKKTAKSYSCOLEFDCY	259
Qy	128	PMEGD-----	BEUKMAFDIETFYHEG--	DEFCKGEIIMISYADEEAEAR	180
Dd	260	APGEYSXMAP	FRILSFDIEACRKGHP	EPHPDVIQI-----ANVTLO	312
Qy	181	-----VDVV--	SNEREMIXRFVQIVREKDP	VDVLITNGDNDFLPYLIKRAE	224
Dd	313	RNVWTLKSCSP	IVGDVMPFETERBVLLAW	RDFIREVDPDIIGYNTCKD	372
Qy	225	KLQVT--LLGRD	KQHP-----EPKIHMGDS	FAVEIKGRIFHDFOLFPV	274
Dd	373	NLXIAEPFIL	GRINSRVAVKDTTFSSRQY	GTRSEKAVEGAVTFDQLQV	432
Qy	275	YTLAEAYEAV	LKTKSLGAEIAIWE-TRES	MKKLAQYSMEDARATYELG	331
Dd	433	YSLNSYSSFL	SEQKEDVHHSIISDLQNG	NAETERRLAVCLDQAYLP	492
Qy	332	EAEALAKLI	QGSV-WDYSRSTGNLVEW	ILLVAVYERNEAPN-KPDB	389
Dd	493	YVEMARVTG	VPISFLLSRQSIKVL-S	QLLARQKNLVI PNAKQSE	546
Qy	390	GYYKEBERG	WEN-ITYLDFRCLYPSI	IVTNVSPDILE-REGCKND	441
Dd	547	ATVLEARAG	FYKPKIATDFASLPSIM	MYNLCTLVIPEARKLNT	606
Qy	442	YKFCCK-a-	FPFGIPSTIGELITMROE	IKKKWKATIDIEKKMLD	499
Dd	607	ETPVKNLQ	KGILPEILLELTARKA	KADLEAKDPLEKAVLDG	666
Qy	500	MGYPKARWY	SKSCASVTAWGHYIEM	TIKIEKP----GF----KVL	551
Dd	667	TGATHIQLP	CLSISSVSYGQMIETHK	VLVEDKFTTLNGYEHNAE	726
Qy	552	PGEKPTIKK	KAEPKYNKLPGLLE	VEGYLGRGFFVAKKRAVI	606
Dd	727	GUSAVEEAM	NLGRGAHSGTFFKPI	KLFEKYYPYLLISKR	786
Qy	607	TRRGLEVVR	DRDSEIAKQAVLEIL	KEDSKEAVEIKVDVVEE	666
Dd	787	DTKGIETVR	DNCLLVQNLVNDCLH	KILIDRIDPGAVQYVKN	846
Qy	667	EQITKDLSEY	KAIQGHVALAKLAKG	IKVRP--GTTISXIVLRG	718
Dd	847	KGLTKTGDD	YEVKAARVELAEWR	KRDAATAPNVGDRP	906
Qy	719	LUSEYDPKH	KHYDDPYIENQVLP	PAVLEILE 749	
Dd	907	VLENNIP----	IDPHYYLENGISKPI	RIFE 933	

RESULT 15  
S75407  
probable DNA-directed DNA polymerase (EC 2.7.7.7) - *Sulfolobus solfataricus*  
N:Alternate names: protein c04041  
C:Species: *Sulfolobus solfataricus*  
C:Date: 11-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change 20-Jun-2000  
C:Accession: S75407  
R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.  
Mol. Microbiol. 22, 175-191, 1996  
A:Title: Organizational Characteristics and information content of an archaeal genome: I  
b:Reference number: S73076; MUID:97055432; PMID:8899719

A: Accession: S75407  
 A: Status: nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-764 <SEN>  
 A: Cross-references: EMBL:Y08257; NID:G1707772; PID:G1707813  
 A: Experimental source: strain P2  
 A: Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996  
 C: Superfamily: herpesvirus DNA-directed DNA polymerase  
 C: Keywords: nucleotidyltransferase

	Query Match	16.8%; Score 676.5; DB 2; Length 764;
	Best Local Similarity	25.3%; Pred.No. 2.6e-27;
	Matches	206; Conservative 168; Mismatches 312; Indels 127; Gaps 24;
Qy	2	IFD'DYITDKGKPIIRIFX-KENGEPKIEDPHFOPVIYALK--DSADDEIKAIKER 58
Dd	7	ILPSEYEIKNTPLVYIWSVDGNSSVLDNNFRPYIIYEGENEIELENIK----- 60
Qy	59	HGKIVRVDAVKKKFLGRDVWVKLIFHFPQDPALRGKIREHPAVIDIYEYDI --- 114
Dd	61	--KNCEALQITKVRYKLGNIVDALLTQTSTPTQIKCKREKISELANNIKGFADIRVTM 118
Qy	115	-----PFA-----KRYLIDKGLIPMEGD-BELKLMAPDIETFYH 147
Dd	119	RYSLDPLDRFTWPAEAVNEVKDFGRTKAYILDKITLSHYGNMPELRITIGVDFQIY-- 177
Qy	148	EGDEFG----KGEEIMISYADEEARVITWKINDLPYVDWNSNEREMIKRFQIVREKD 202
Dd	177	--SKYGSLNPKRQPIVVMSLWSKEGPMQFS-----LDEGIDDLKIIRFVDYILNYD 226
Qy	203	PQVLITTINGDNFDLPYLIKRAEKLVTLGLGRDKEHPKPIHRMGDSFAY-----BIKG 256
Dd	227	PDIIFVSDLLPWKYITERASSLGWKIDIGR-----KIGSEVSQGTGYHSISG 276
Qy	257	RIHFDPVPVVRRTINPTYTLEAVEZAVLGKTCKSLGAE--ETAATAWETEESMKLAQYS 314
Dd	277	RLAVDLTGILLNERSLGHVDLIDV-SNYLGISFSRYSFKWYEISRYWDNEKNRIIREYS 335
Qy	315	MEDARATYELGKEFFPWMEABRLAKIGOSVWDVRSSTGNLVEVYLLRVAVERNELAPNKP 374
Dd	336	IENARSYYLLGNLYLSTYSBLVINGVLPLDKLSVASWGNRIETSLIRTATKSGELIPIRM 395
Qy	375	DEEYRRRLRTTVLGGVKEPERGLMENTYLFRCILYPSLIIVTHNVSPDTLREGCNKY 434
Dd	396	DNPNRSKIKK---NLIQPKVGITYDVTYLDISSVYSLVIRKFNIAPTLLVKEQCDDC 451
Qy	435	DVAPIGVYKCFDPPGISLIGELITMRQEIKKKMKATIDPLETKOMLDYRPQAVKLHAN 494
Dd	452	YSSPISNYKFKREPSPGLYKTFDELSELNVRDSNKIV-----IEELI-----S 493
Qy	495	SYGYNGYPCARYKCEASVTAGSHVIEMTIKIEBKFGPVLYADTDGYATIPEG 554
Dd	494	SFNDYVHVWNARYSREIASAFDEFNEIIRPIDLIKGS-GUDVLANDLLLTFTV---- 548
Qy	555	KPETIKKKAKEFLUKYINSKLPGLLELEYEGFYLRGFFVAKRVAVIDEGRITTRGLEVV 614
Dd	549	--GGSRDKNVELITKNSLVN--LDVKVKIFYKSLLVLDNNRYAGLSEGDKI-----DIA 599
Qy	615	RR-----DWSEIAKETCAKULEALKEDSVKAVEIVKQVBEIAKYQVPLEKLVIEHQI 659
Dd	600	RKGEEDMWLCELARNIKRKIIIBEILTSKOVKKAIKLVKSTVTKLRERGEFDNEELITWAKI 659
Qy	670	TCDLSYKAIQHPVAIXELAAGIKVRPGTITSYIVLRSGSGKISRQV---ILLSEYDPK 726
Dd	660	ERDLNEVNNCLPVTAAARKAISQGYLISKDSKIGYIVVKGLGPLNDRAEPFFLVKE---- 715
Qy	727	XHKYDPPDYIENQVLPVAVRILEAFQYRKEDLK 759
Dd	716	KNRIDIEYVD-QIFRETIKLKLPLGVNEESLK 747

Search completed: February 9, 2004, 11:53:19  
Job time : 61 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2004, 11:39:55 ; Search time 23 Seconds  
(without alignments)  
1582.551 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026

Sequence: 1 MFEDTYTKDGRPIIRIFK.....KEDLYQSSKQVGLDAWLKX 774

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3446.5	85.6	1829	1 DPOL_THEST	Q33845 thermococcu
2	3356	83.4	1702	1 DPOL_THELI	P30317 thermococcu
3	3253.5	80.8	773	1 DPOL_THESO	P56689 thermococcu
4	3204.5	79.6	775	1 DPOL_THES9	Q56366 thermococcu
5	3164.5	78.6	771	1 DPOL_PYRAB	P79161 pyrococcus
6	3151	78.3	775	1 DPOL_PYRFU	P80061 pyrococcus
7	2934.5	72.9	1312	1 DPOL_PYRSD	Q51334 pyrococcus
8	2886	71.7	1235	1 DPOL_PYRHO	Q59610 pyrococcus
9	2809	69.8	1523	1 DPOL_PYRHM	P74918 thermococcu
10	2782	69.1	1671	1 DPOL_PYRKO	P77933 pyrococcus
11	2769.5	68.8	1699	1 DPOL_THRGS	Q9hh84 thermococcu
12	2535.5	63.0	1668	1 DPOL_THERY	Q9hh05 thermococcu
13	1392	34.6	824	1 DPOL_METVO	P52025 methanococc
14	1301	32.3	781	1 DPOL_ARCFU	Q29753 archaeoglob
15	1292	32.1	1634	1 DPOL_METUA	Q58235 methanococc
16	1209.5	30.0	586	1 DPOL_METTH	Q27276 methanobact
17	1138	28.3	784	1 DPOL_AERPE	Q93746 aeropyrum p
18	697.5	17.3	1105	1 DPOL_ORYSA	Q91r66 oryza sativ
19	693.5	17.2	1088	1 DPOL_SOVEN	Q48901 glycine max
20	691	17.2	763	1 DPOL_SULSH	O05706 sulfolobus
21	676.5	16.8	764	1 DPOL_SULSO	P95979 sulfolobus
22	676.5	16.8	1081	1 DPOL_ARATH	Q31vnt arabidopsis
23	657.5	16.3	1038	1 DPOL_CANAL	P46588 candida alb
24	657.5	16.3	1086	1 DPOL_SCHPO	P30316 schizosacch
25	624.5	15.5	1092	1 DPOL_DROME	P54358 drosophila
26	621.5	15.4	1097	1 DPOL_YEAST	P15436 saccharomyc
27	619.5	15.4	1103	1 DPOL_RAT	Q54747 rattus norv
28	616.5	15.3	872	1 DPOL_SULOH	O50607 sulfuriapha
29	616.5	15.3	882	1 DPOL_SULSO	P26811 sulfolobus
30	616.5	15.3	1105	1 DPOL_MOUSE	P52431 mus musculus
31	615.5	15.3	1103	1 DPOL_MESAU	P97283 mesocricetu
32	614.5	15.3	1094	1 DPOL_PLAFK	P20315 plasmodium
33	611	15.2	1107	1 DPOL_HUMAN	P28340 homo sapien

34	609	15.1	1513	1 DPOA_OXYTR	Q27152 oxytricha t
35	605.5	15.0	1106	1 DPOL_BOVIN	P28339 bos taurus
36	587	14.6	1492	1 DPOA_OXYNO	Q94636 oxytricha n
37	584	14.5	1451	1 DPOA_RAT	O89042 rattus norv
38	583.5	14.5	1462	1 DPOA_HUMAN	P99884 homo sapien
39	581	14.4	1081	1 DPOL_CAEEL	P90829 caenorhabdi
40	578.5	14.4	875	1 DPOL_SULAC	P95690 sulfolobus
41	570.5	14.2	959	1 DPOL_AERPE	O93745 aeropyrum p
42	570	14.2	1465	1 DPOA_MOUSE	P33609 mus musculu
43	569.5	14.1	1243	1 DPOA_ORYSA	O48653 oryza sativ
44	548.5	13.6	1339	1 DPOA_TRYBB	P27727 trypanosoma
45	527.5	13.1	1468	1 DPOA_YEAST	P13382 saccharomyc

## ALIGNMENTS

### RESULT 1

ID	DPOL_THEST	STANDARD;	PRT; 1829 AA.
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	DNA polymerase (EC 2.7.7.7).		
GN	POL.		
OS	Thermococcus sp. (strain TY).		
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;		
OX	Thermococcus.		
NCBI	TaxID=110163;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98094267; PubMed=9434178;		
RA	Niehaus F., Frey B., Antranikian G.;		
RT	"Cloning and characterisation of a thermostable alpha-DNA polymerase		
RT	from the hyperthermophilic archaeon Thermococcus sp. TY."		
RL	Gene 204:153-158(1997).		
CC	!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate		
CC	+ [DNA](N).		
CC	!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES		
CC	A POST-TRANSLATIONAL EXCISION OF THE THREE INTERVENING REGION		
CC	(INTERNS) FOLLOWED BY PEPTIDE LIGATION.		
CC	!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL; Y13030; CAA73475.1; ..		
DR	HSSP; P56689; ITGO		
DR	InterPro; IPR006172; DNA_pol_B.		
DR	InterPro; IPR006134; DNA_pol_B_dom.		
DR	InterPro; IPR006133; DNA_pol_B_exo.		
DR	InterPro; IPR003586; Hedgehog_HintC.		
DR	InterPro; IPR003587; Hedgehog_hintN.		
DR	InterPro; IPR006141; Intein.		
DR	InterPro; IPR006142; INTEIN.		
DR	InterPro; IPR004042; Intein_endonuc.		
DR	InterPro; IPR004578; Pol2.		
DR	Pfam; PF00136; DNA_pol_B; 4.		
DR	Pfam; PF03104; DNA_pol_B_exo; 1.		
DR	PRINTS; PR00379; INTEIN.		
DR	SMART; SM00305; HintC; 3.		
DR	SMART; SM00305; HintN; 3.		
DR	SMART; SM00486; POLSC; 1.		
DR	TIGRFAMs; TIGR01443; Intein_Cterm; 3.		
DR	TIGRFAMs; TIGR01445; Intein_Nterm; 3.		
DR	TIGRFAMs; TIGR00592; pol2; 2.		
DR	PROSITE; PS00116; DNA_POLYMERASE_B; FALSE_NEG.		

DR PROSITE; PS50818; INTEIN\_C\_TER; 3.  
DR PROSITE; PS50819; INTEIN\_ENDONUCLEASE; 2.  
DR PROSITE; PS50817; INTEIN\_N\_TER; 3.  
KW Transferrase; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;  
KW Protein splicing.  
FT CHAIN 1 409 DNA POLYMERASE, 1ST PART (POTENTIAL).  
FT CHAIN 410 769 INTEIN I.  
FT CHAIN 770 855 DNA POLYMERASE, 2ND PART (POTENTIAL).  
FT CHAIN 856 1392 INTEIN II.  
FT CHAIN 1393 1441 DNA POLYMERASE, 3RD PART (POTENTIAL).  
FT CHAIN 1442 1598 INTEIN III.  
FT CHAIN 1599 1829 DNA POLYMERASE, 4TH PART (POTENTIAL).  
SQ SEQUENCE 1829 AA; 211875 MW; A113A8BC57EB9CB3 CRC64;  
  
Query Match 85.6%; Score 3446.5; DB 1; Length 1829;  
Best Local Similarity 42.2%; Pred. No. 1.2e-179;  
Matches 771; Conservative 0; Mismatches 3; Indels 1055; Gaps 3;  
  
QY 1 MIFPDYITKDGKPIIRIFKXENGEPKIELDPHPQPIYIALLKDDSAIDIKAIKGERHG 60  
DB 1 MILDYITKDGKPIIRIFKXENGEPKIELDPHPQPIYIALLKDDSAIDIKAIKGERHG 60  
  
QY 61 KIVRVDAVKKKPLGRDVEVKLIPEHPDVPALRGKIREHPAVIDIYEYDIPPAKRY 120  
DB 61 KIVRVDAVKKKPLGRDVEVKLIPEHPDVPALRGKIREHPAVIDIYEYDIPPAKRY 120  
  
QY 121 LIDKGLIPMEGDEELKMAFDIETFYHEGDSFGKEIIMISYADEEARVITWKNIDL 180  
DB 121 LIDKGLIPMEGDEELKMAFDIETFYHEGDSFGKEIIMISYADEEARVITWKNIDL 180  
  
QY 181 VDVVSNREMIKRFVQIVREKDPDVLITNGDNLPLVLIKRAEKIGVTLLGLGRDKEH 240  
DB 181 VDVVSNREMIKRFVQIVREKDPDVLITNGDNLPLVLIKRAEKIGVTLLGLGRDKEH 240  
  
QY 241 PKIHRMGDSFAVEIKGRIFHDLFPVVRTINLPTTYLEAVYEAVLGKTSKLGAEIAA 300  
DB 241 PKIHRMGDSFAVEIKGRIFHDLFPVVRTINLPTTYLEAVYEAVLGKTSKLGAEIAA 300  
  
QY 301 WETESMKLAQYSMEDARATYELGKEFFPMEAKLIGOSVMDVSRSSTGNLVEWYLL 360  
DB 301 WETESMKLAQYSMEDARATYELGKEFFPMEAKLIGOSVMDVSRSSTGNLVEWYLL 360  
  
QY 361 RVAYERNEAPNKPDEEYRRRLRTTYLGGYVKEPERGLWENITVLDPRC 410  
DB 361 RVAYERNEAPNKPDEEYRRRLRTTYLGGYVKEPERGLWENITVLDPRC 410  
  
QY 411 ----- 410  
DB 421 GKGIVNISDVKEGDYILGIGWQVRKVKYHYEGKLININGLKCTPNHKVPVVTENDRQ 480  
QY 411 ----- 410  
DB 481 TRIRDSLAKSFLSGKVKGKIIITTKLFEKIAEPKKNPSEEBILKLGELSGIILAEGTLRK 540  
QY 411 ----- 410  
DB 541 DIEYFDSRSGKKRISHQYRVEITIGENEKELLERYIFDKLFGIRPSYKXKGDNALKI 600  
QY 411 ----- 410  
DB 601 TTAKKAVYLOIEELLKNIESLYAPAVLRGFFPERDATVKNIRSTIVVTOGNNKWKIDIVA 660  
QY 411 ----- 410  
DB 661 KLKDSLGIPYSRVEYKYIENGKELTKHILEITGRDGLILFOTLVGFISSKNEALEKAIE 720  
QY 411 ----- -LYPSIIVTHN 420  
DB 721 VREWNRLKNSFNLSSTFEVSSEYKGEVYDITLGNPPYFANGILTHNSLYPSIIVTHN 780  
QY 421 VSPDTLEREGCKNYDVAPIGVYKFCDFPGFIPSGILGELITMROEIKKKMKATIDPIEKK 480  
|||||

RESULT 2  
DPOL\_THELI

DB 781 VSPDTLEREGCKNYDVAPIGVYKFCDFPGFIPSGILGELITMROEIKKKMKATIDPIEKK 840  
QY 481 MLDYRQRAVKLHA ----- 493  
DB 841 MLDYRQRAVKLLANSILPNEWLPPIENGVEKVFVIGFIDRYMEEQDKVTVTDNTEVLE 900  
QY 494 ----- 493  
DB 901 VDNIFAFSLNKESKSEIKVKALIRHKYGEAYEVELNSGRKIITRGHSLFTIRNGKI 960  
QY 494 ----- 493  
DB 961 KEIMGEEVKVGDLIIIVPKVKVNEKEAVINIPELISKLPDEDTADVMTTPVKGKKNPFK 1020  
QY 494 ----- 493  
DB 1021 GMLRTLKWIIFGEBESKRIRTFNRYLFLHEELGFVKLLPRGYEVTDWEGLRKYRQLYEKLVK 1080  
QY 494 ----- 493  
DB 1081 NLRYNGNKREYLVRFNDIKDSVCPFRKELBEWKITGKPRXKCILKVDSDPGFLGYI 1140  
QY 494 ----- 493  
DB 1141 VSEGYAGAQNKTGCMYSYVKLYNENPNVLKMNIAEKFVKVAVGKNQVCDIPKQWYL 1200  
QY 494 ----- 493  
DB 1201 LAKSLCGVTAENKRIPSIIFDSSEPRWAFURAYFVGDDIHPSKRLSLSTKSELLANQL 1260  
QY 494 ----- 493  
DB 1261 VFLLNSLGVSSIKIGFDSGVYRVVINEDLPFLQTSRQNTVYPNLIPEKVEEIPGRAFK 1320  
QY 494 ----- 493  
DB 1321 KNITPEKPELADSGKDKRKVKLLDFLLNGDIVLDVKNVKEKRYEGVYDLSVEDNEN 1380  
QY 494 ----- NSYCYGMYGPKARWYKSCAESVTAWGRHYIEMTIKEEKFGEKVLAYA 542  
DB 1381 FLVGFGLLVAHNSYGYMGYPKARWYKSCAESVTAWGRHYIEMTIKEEKFGEKVLAYA 1440  
QY 543 D ----- 543  
DB 1441 DSVTGDTEIIVKRNRIEFVPIEKLFRVYRIGEKEYCILBDVEALTLDNRGKLIWKV 1500  
QY 544 ----- 543  
DB 1501 PYWRHRAKKVYRIWITNSWYIDYTDHSLIVAEDGLKEARPMEIEGKSLIATKDDLSG 1560  
QY 544 ----- TDGFIATIPGEKPTIKKAKE 565  
DB 1561 VEYIKPHAIEISYNGYVYDIEVEGTHRFPFANGILVHNTDGFYATIPGEKPTIKKAKE 1620  
QY 566 FLKVIKNSKLPGLLELEVEGFYLRGPFVAKRYAVIDEGRITTRGLEVVVRDSEIASEKT 625  
DB 1621 FLKVIKNSKLPGLLELEVEGFYLRGPFVAKRYAVIDEGRITTRGLEVVVRDSEIASEKT 1680  
QY 626 QAKVLEAILKEDSVEKAVEIKVWVEEIAKYQVPEKLVHQBQITKOLSEYKAIGPHVAI 685  
DB 1681 QAKVLEAILKEDSVEKAVEIKVWVEEIAKYQVPEKLVHQBQITKOLSEYKAIGPHVAI 1740  
QY 686 AKELAAKGIKVRPGTTIISYIVLRGSKISDRVILLSEYDPKHKYDPPDYIENQVLPVAVL 745  
DB 1741 AKELAAKGIKVRPGTTIISYIVLRGSKISDRVILLSEYDPKHKYDPPDYIENQVLPVAVL 1800  
QY 746 RILEAFGYRKEDLYQSSKQVGLDAMLKX 774  
DB 1801 RILEAFGYRKEDLYQSSKQVGLDAMLKX 1829

ID DPOL\_THELI STANDARD; PRT; 1702 AA.  
AC P30317;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA polymerase (EC 2.7.7.7) (Vint DNA polymerase) [Contains:  
DE Endonuclease PI-Tlii (EC 3.1.-.-) (Tli pol-1 intein) (IVPS2);  
DE Endonuclease PI-Tlii (EC 3.1.-.-) (Tli pol-2 intein) (IVPS1)].  
GN POL.  
OS Thermococcus litoralis.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Thermococcus.  
OX NCBI\_TaxID=2265;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=92302285; PubMed=1608969;  
RA Perler F.B., Comb D.G., Jack W.E., Moran L.S., Qiang B.,  
RA Kucera R.B., Benner J., Slatko B.E., Nwankwo D.O., Hempstead S.K.,  
RA Carlow C.K.S., Jannasch H.;  
RT "Intervening sequences in an Archaea DNA polymerase gene.";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:5577-5581(1992).  
RN [2]  
RP PROTEIN SPLICING.  
RX MEDLINE=93117083; PubMed=1475179;  
RA Hodges R.A., Perler F.B., Noren C.J., Jack W.E.;  
RT "Protein splicing removes intervening sequences in an archaea DNA  
RT polymerase.";  
RT Nucleic Acids Res. 20:6153-6157(1992).  
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
CC EXHIBITS 3', TO 5', EXONUCLEASE ACTIVITY.  
CC -1- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE  
CC INTEIN MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY  
CC ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE  
CC INTEIN.  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA}(N).  
CC -1- PFM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
CC A POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION  
CC (INTEINS) FOLLOWED BY PEPTIDE LIGATION.  
CC -1- BIOTECHNOLOGY: Used in the PCR method because of its high  
CC thermostability and low error rate. Sold by New England Biolabs.  
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING  
CC ENDONUCLEASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M74198; AAA72100.1; -;  
CC EMBL; M74198; AAA72101.1; -;  
CC DR PIR; S42459; S42459.  
CC DR HSP; P56689; ITGO.  
CC DR REBASE; 2621; PI-Tlii.  
CC DR REBASE; 2621; PI-Tlii.  
CC DR InterPro; IPR006172; DNA\_pol\_B.  
CC DR InterPro; IPR006134; DNA\_pol\_B\_dom.  
CC DR InterPro; IPR006133; DNA\_pol\_B\_exo.  
CC DR InterPro; IPR003586; Hedgehog\_hintC.  
CC DR InterPro; IPR003587; Hedgehog\_hintN.  
CC DR InterPro; IPR006141; intein.  
CC DR InterPro; IPR006142; intein.  
CC DR InterPro; IPR004042; intein\_endonuc.  
CC DR InterPro; IPR004578; Pol2.  
CC DR Pfam; PF00136; DNA\_pol\_B; 3.  
CC DR Pfam; PF00104; DNA\_pol\_B\_exo; 1.  
CC DR PRINTS; PR00379; INTEIN.  
CC DR SMART; SMO0305; HintC; 2.  
CC DR SMART; SMO0306; HintN; 2.

DR SMART; SMO0486; POLBc; 1.  
DR TIGRFAMS; TIGR01443; intein\_Cterm; 2.  
DR TIGRFAMS; TIGR01445; intein\_Nterm; 2.  
DR TIGRFAMS; TIGR00592; pol2; 2.  
DR PROSITE; PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.  
DR PROSITE; PS00818; INTEIN\_C\_TER; 2.  
DR PROSITE; PS00819; INTEIN\_ENDONUCLEASE; 2.  
DR PROSITE; PS00817; INTEIN\_N\_TER; 2.  
KW Transferase; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;  
KW Protein splicing; Intrin homing.  
FT CHAIN 1 494 DNA POLYMERASE, 1ST PART.  
FT CHAIN 495 1032 ENDONUCLEASE PI-Tli I.  
FT CHAIN 1033 1081 DNA POLYMERASE, 2ND PART.  
FT CHAIN 1082 1471 ENDONUCLEASE PI-Tli I.  
FT CHAIN 1472 1702 DNA POLYMERASE, 3RD PART.  
SQ SEQUENCE 1702 AA; 197293 MW; 21D6B98C75F53B20 CRC64;  
Query Match 83.4%; Score 3356; DB 1; Length 1702;  
Best Local Similarity 42.7%; Pred. No. 8.7e-175;  
Matches 726; Conservative 33; Mismatches 15; Indels 928; Gaps 2;  
QY 1 MIFDTYITKDGKPIIRIFKKEGFEKIELDPHPQPIYVALLKDDSAIDRIKAIKGRHG 60  
DB 1 MILDTYITKDGKPIIRIFKKEGFEKIELDPHPQPIYVALLKDDSAIEIKAIKGRHG 60  
QY 61 KIVVDAVKVKKFKLGRDVEVWVKLIEHPQDVDPALRGKIREHPAVIDIYVDIPFARKY 120  
DB 61 KTVRLDAVKRKFKLGRDVEVWVKLIEHPQDVDPAMRGKIREHPAVIDIYVDIPFARKY 120  
QY 121 LIDKGLIPMEGDEBELKMAPDIETFYHEGDFGKEIIMISYADEEERARVITWKNIDLPY 180  
DB 121 LIDKGLIPMEGDEBELKMAPDIETFYHEGDFGKEIIMISYADEEERARVITWKNIDLPY 180  
QY 181 VDVVSNREMIKRFVQVREKDPDLITYGNDPDLPIYIKRAEKLGVTLGLGRDKEHPE 240  
DB 181 VDVVSNREMIKRFVQVREKDPDLITYGNDPDLPIYIKRAEKLGVTLGLGRDKEHPE 240  
QY 241 PKIHRMGDSFAVEIKGRIHFDLPVVRTTNLPYTLVAVVEAVLGTGKSGAEETAAI 300  
DB 241 PKIHRMGDSFAVEIKGRIHFDLPVVRTTNLPYTLVAVVEAVLGTGKSGAEETAAI 300  
QY 301 WETESMKKLAQYSMEDARATYELGKFPFMEABLAKLIGQSVMDVSRSTGNLVEWYLL 360  
DB 301 WETESMKKLAQYSMEDARATYELGKFPFMEABLAKLIGQSVMDVSRSTGNLVEWYLL 360  
QY 361 RVAVERNELAPNKDESEYRRRLTTLVGGVKEPEGLWENITLDFCLYPSIIVTHN 420  
DB 361 RVAVARNELAPNKDESEYRRRLTTLVGGVKEPEGLWENITLDFCLYPSIIVTHN 420  
QY 421 VSPDTLEREGCKNDVAPIVGYKFCDFPGFIPSIIGELITMRQEIKKQKATIDPIEKK 480  
DB 421 VSPDLEKEGCKNDVAPIVGYRCKDFPGFIPSIIGDLIARQDIKKQKSTIDPIEKK 480  
QY 481 MLDYRQRAVKLHA----- 493  
DB 481 MLDYRQRAIKLLANSILPNEWLPPIENGEEKFKVIGFINSYMEKQKENVKTVTEVLE 540  
QY 494----- 493  
DB 541 VNNLFATSFNKKIKESVKKVKKALIRHKYKGAKEIQLSSGRKINITAGHSLFTVRNGEI 600  
QY 494----- 493  
DB 601 KEVSGDGIKEGDLIVAPKKIKLNEKGVSNIPELISDLSEETADIVMTISAKGRKNFFK 660  
QY 494----- 493  
DB 661 GMLATLRWFMGEENRIRTFNRYLFHLEKGLIKLPLRGYEVTDWELKKYKQLYEKLAG 720  
QY 494----- 493  
DB 721 SVKYNKNREYLVNMFNBKIDFISYFPOKELEEWKIGTNGFRNTCNILKVDDEDFGLAGY 780

QY 494 ----- 493  
Db 781 VSEGAGAKNKTGIGSYVKLYNEDPNVLSMKNVAEKFFGKVRVDNCVSIKGMAYL 840  
QY 494 ----- 493  
Db 841 VMKCLGALAENKRIPSVILTSPEVRNFSLEAYFTGDDIHPSKRFSLSTKSELLANQL 900  
QY 494 ----- 493  
Db 901 VFLNSLGISSVKGIGDSGVVRVYNEDLPQFQTSRKNNTYNSLNPKRILRDVFGKEFQ 960  
QY 494 ----- 493  
Db 961 KMTFKFKELVDSKLNREKAKLEFFTINGDIVLDVYKSKEDYEGVYVLDLSDVEDNEN 1020  
QY 494 ----- 543  
Db 1021 FLVGFGLLYAHNSYGYGPKARWYSKCAESVTANGRHVYEMTIRIEEKFGFKVLYA 1080  
QY 543 D----- 543  
Db 1081 DSVSGESLIIIRQNKIRPKIKDLPSKVDYSIGKEYCILEGVEALTLDDDGKLVKVPV 1140  
QY 544 ----- 543  
Db 1141 PYVHRHRANKRMFRIWLTNSWYIDVTEDHSLIGLNTSKTKAKKIGERLKEVKVPFELGK 1200  
QY 544 ----- 543  
Db 1201 AVKSLICNPAPLKDENTKTSBIAVFWELVGLVGDGNGWGDGSRWAETVYLGSLGKQABE 1260  
QY 544 ----- 543  
Db 1261 IKQKLEPLKTYGVISNYYPKNEKGFNLLAKSLVKFMKRFKDEKGRRKIPFMYELPV 1320  
QY 544 ----- 543  
Db 1321 TYIEAFLRGLFSADGTVTIRKGVPEIRLTNIDADFLREVRLKLWIVGINSIFAETPNR 1380  
QY 544 ----- 543  
Db 1381 YNGVSTGYSKHLRIKKNRFAERIGFLIERKQKLEHLKSARVKRNTIDFGDLVHVK 1440  
QY 544 ----- 572  
Db 1441 KYEPIPYGVYVDIEVEETHRPFANNILVHNTDGFYATIPGKPELIIKKAKEFLNYNS 1500  
QY 573 KLPGLLELEYEGFYLRGFFVAKRYAVIDEGRIITRGLVVRDMSIAKETQAKVLEA 632  
Db 1501 KLPGLLELEYEGFYLRGFFVTKRYAVIDEGRIITRGLVVRDMSIAKETQAKVLEA 1560  
QY 633 ILKEDSVEKAVIVKDVVEEIAKYQVPLEKLVHIOITKDLSEYKAIQPHVAIAKRLAAK 692  
Db 1561 ILKEGSEKAVEVRDWEKIAKYRVPLEKLVHIOITKDLSEYKAIQPHVAIAKRLAAK 1620  
QY 693 GIKVRPTIISIVLRGSKISDRVILLSEYQPKHKYDPDYVYIENQVLPVLRILEAFG 752  
Db 1621 GIKVKEPTIISIVLRGSKISDRVILLTEYDPRKHKYPDYVYIENQVLPVLRILEAFG 1680  
QY 753 YRKEDLKYSQSKQVGLDAILKK 774  
Db 1681 YRKEDLYQSSKQTGLDAWLKR 1702

## RESULT 3

DPOL\_THEGO

ID DPOL\_THEGO STANDARD; PRT; 773 AA.

AC P56689;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA polymerase (EC 2.7.7.7) (TO POL).  
GN POL OR POLA.  
OS Thermococcus gorgonarius.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcaceae;  
OC Thermococcus.  
OX NCBI\_TaxID=71997;  
RN [1]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=99199230; PubMed=10097083.  
RA Hopfinger K.P., Eichinger A., Engh R.A., Laue F., Ankenbauer W.,  
RA Huber R., Angerer B.;  
RT "Crystal structure of a thermostable type B DNA polymerase from  
RT Thermococcus gorgonarius";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:3600-3605(1999).  
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
CC EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA}(N).  
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
CC PDB; 1TGO; 22-MAR-99.  
DR InterPro: IPR006172; DNA\_pol\_B.  
DR InterPro: IPR006134; DNA\_pol\_B\_dom.  
DR InterPro: IPR006133; DNA\_pol\_B\_exo.  
DR InterPro: IPR004578; pol2.  
DR Pfam: PF00136; DNA\_pol\_B\_1.  
DR Pfam: PF03104; DNA\_pol\_B\_exo; 1.  
DR PRINTS; PR00106; DNAPOLB.  
DR SMART; SM00486; POLBc; 1.  
DR TIGRFAMs; TIGR00592; pol2; 1.  
DR PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
KW Transferase; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding; Hydrolase; Nuclease; Exonuclease; Multifunctional enzyme;  
KW 3D-structure.  
FT DISULFID 428 442  
FT STRAND 2 10  
FT TURN 11 12  
FT STRAND 13 22  
FT TURN 23 24  
FT STRAND 25 31  
FT STRAND 37 42  
FT HELIX 45 47  
FT HELIX 48 51  
FT TURN 52 53  
FT STRAND 55 58  
FT TURN 59 60  
FT STRAND 61 64  
FT STRAND 67 75  
FT TURN 76 77  
FT STRAND 78 86  
FT TURN 90 91  
FT HELIX 92 102  
FT TURN 104 105  
FT STRAND 106 110  
FT HELIX 116 123  
FT TURN 124 125  
FT STRAND 137 144  
FT TURN 152 153  
FT STRAND 157 164  
FT TURN 165 166  
FT STRAND 167 172  
FT TURN 179 180  
FT STRAND 181 183  
FT HELIX 187 201  
FT STRAND 205 208  
FT HELIX 211 213  
FT TURN 214 214  
FT TURN 215 225  
FT TURN 226 227  
FT TURN 234 235  
FT STRAND 240 243  
FT STRAND 248 251  
FT TURN 253 254

FT	STRAND	256	259
FT	HELIX	260	267
FT	HELIX	275	283
FT	HELIX	292	301
FT	TURN	303	304
FT	HELIX	305	337
FT	TURN	338	338
FT	HELIX	341	345
FT	TURN	346	346
FT	HELIX	349	363
FT	TURN	364	365
FT	STRAND	366	366
FT	HELIX	374	379
FT	STRAND	389	390
FT	STRAND	396	405
FT	HELIX	408	415
FT	TURN	416	417
FT	TURN	420	422
FT	STRAND	423	423
FT	TURN	424	424
FT	TURN	426	427
FT	STRAND	431	433
FT	TURN	435	437
FT	STRAND	440	442
FT	HELIX	448	469
FT	HELIX	473	490
FT	TURN	491	492
FT	HELIX	493	498
FT	TURN	500	501
FT	TURN	503	504
FT	HELIX	507	530
FT	TURN	531	532
FT	STRAND	535	540
FT	STRAND	543	547
FT	TURN	549	550
FT	HELIX	553	568
FT	TURN	569	570
FT	TURN	573	574
FT	STRAND	577	590
FT	TURN	591	592
FT	STRAND	593	597
FT	TURN	599	600
FT	STRAND	603	606
FT	STRAND	615	615
FT	HELIX	617	631
FT	TURN	632	633
FT	HELIX	636	651
FT	TURN	652	653
FT	STRAND	657	659
FT	STRAND	662	665
FT	HELIX	670	672
FT	HELIX	679	689
FT	TURN	690	690
FT	TURN	695	696
FT	STRAND	698	704
FT	TURN	710	711
FT	STRAND	714	716
FT	HELIX	717	719
FT	TURN	722	724
FT	STRAND	727	727
FT	HELIX	729	735
FT	TURN	736	737
FT	HELIX	738	740
FT	TURN	741	741
FT	HELIX	742	746
FT	TURN	747	749
FT	HELIX	752	754
FT	TURN	766	767
FT	TURN	771	773
SQ	SEQUENCE	773 AA; F67AF04E875FBE44 CRC64;	

Query Match 80.8%; Score 3253.5; DB 1; Length 773;  
Best local similarity 78.3%; Pred. No. 1.2e-169;

Matches 605; Conservative 83; Mismatches 82; Indels 3; Gaps 2;			
QY	1	MIFDTDYITKOGKPIIRIFKKEGEBFKIELDPHFQPIYIALLKODSAIDEIKAIKGERHG	60
DB	1	MILDTDYITEDGKPVIRIFKKEGEBFKIDYDRNFEPYIYALLKODSAIEDVKKITAERHG	60
QY	61	KIVRVVDVAVKVKKFLGRDVEVWKLIIEHPQDVVALRGKIREHPAVDIIYDYDIPFAKRY	120
DB	61	TTVRVVRRAEKVKKFLGRPIEVWKLIYFTHPDVFAIRDKIKEHPAVDIIYDYDIPFAKRY	120
QY	121	LIDKGLIPMEGDEELKMAFDIETFYHGEDFSGKEIIMISYADEEERAVITWKNIDILPY	180
DB	121	LIDKGLIPMEGDEELKMAFDIETLYHEGEFAFGPILMISYADEGARVITWKNIDILPY	180
QY	181	VDVVSNEREMIKRFVQIVREKDPDVLITYANGDNFDLPYLKRAKLGITVLLGRDKHEPE	240
DB	181	VDVVSNEKEMIKRFLKVVKEKDPDLITYNGDNFDLPYLKRAKLGITVLLGRDKHEPE	240
QY	241	PKIHRMGDSFAVEIKGRIHFDLPFVVRRTINLPYTTLEAVVEAVLGTKSKLGAEEIAAI	300
DB	239	PKIQRMGDRFAVEIKGRIHFDLPVIRRTINLPYTTLEAVVEAIFGQPKKVAEEIAQA	298
QY	301	WETBESMKKLAQYSMEDARATYELGKEFPFMEASLAKLIGQSVWDVRSSTGNJVEWVLL	360
DB	299	WETGEGLEVARYSMEDAKVTYELGKEFPFMEAGLSRLVGQSLWDVRSSTGNJVEWVLL	358
QY	361	RVAYERNELAPNKPDEEYRRRLRTTYLGGYVVKPERGLWENITYDPRCLYPSIIVTHN	420
DB	359	RKAYERNELAPNKPDERELAE-RESYAGGYVVKPERGLWENITVDFRSIYPSIIVTHN	417
QY	421	VSPOTLEREGCKNDVAPIVGYKFCCKPPGPHIPILOBLITMRQBIKKQKMATIDPIEKK	480
DB	418	VSPDTLNRGCEYDVAPQVGHKFCCKPPGPIPSLLGLDLEERQKVKQKMATIDPIEKK	477
QY	481	MLDYRQRAVKLHANSYGYMGYPKARWYSKEASVTAWGRHYTEMITKEEERKFGKVL	540
DB	478	LLDYRQRAIKILANSFYGYGYAKARWYCKEASVTAWGRQYIETITREIEERKFGKVL	537
QY	541	YADTDGFYATIPGKPKETIKKAKKFLKYNKSLUGLLELEYEGFYLRGPFVAKRYAVI	600
DB	538	YADTDGFFATIPGADAETVKKAKKFLDYINAKLPGLELEYEGFYKRGPFVTKKRYAVI	597
QY	601	DEGRITTRGLEVVRRDWESEIAKETQAKLEAILKEDSVKAEIVKDVVEEIAKYOVPL	660
DB	598	DEEDKITTRGLEIVRRDWESEIAKETQARVLEAILKHGDVEAEVRIKVEKESKYEVPP	657
QY	661	EKLVIHQITKDLSEYKAIGPHVAIAKRLAAKGKVRPGTIIISYIVLRGSKISDRVILL	720
DB	658	EKLVIHQITRDLKDYKATGPHVAVAKRLAARGIKIRPGTIVISIVLRGSGRIGDRAIPP	717
QY	721	SEYDPKXKHYDPDYVIENQVLPVLRILEAFGRKEDLKYQSSKQVGLDAWLK	773
DB	718	DEFDPAKHKYDAEYIIENQVLPVLRILEAFGRKEDLRYQKTRQVGLGAWLK	770
RESULT 4			
DPOL_THES9 STANDARD; PRT; 775 AA.			
AC	O56366;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, last sequence update)		
DT	15-SEP-2003 (Rel. 42, last annotation update)		
DE	DNA polymerase (EC 2.7.7.7).		
GN	POL OR POLA.		
OS	Thermococcus sp. (strain 90N-7).		
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;		
OC	Thermococcus.		
OX	NCBI_TaxID=103799;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96224274; PubMed=8643567;		
RA	Southworth M.W., Kong H., Kucera R.B., Ware J., Jannasch H.W.,		
RA	Perler F.B.,		

"Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on Thermococcus sp. 90N-7 and mutations affecting 3'-5' exonuclease activity.";  
 Proc. Natl. Acad. Sci. U.S.A. 93:5281-5285 (1996).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + (DNA)(N)  
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U47108; AAA88769.1; -  
 CC PIR; S67920; S67920.  
 CC PDB; 1QHT; 04-JUN-00.  
 CC InterPro; IPR006172; DNA\_pol\_B.  
 CC InterPro; IPR006134; DNA\_pol\_B\_dom.  
 CC InterPro; IPR006133; DNA\_pol\_B\_exo.  
 CC InterPro; IPR004578; Pol2.  
 CC Pfam; PF00136; DNA\_pol\_B\_1.  
 CC Pfam; PF03104; DNA\_pol\_B\_exo; 1.  
 CC PRINTS; PR00106; DNAPOLE.  
 CC SMART; SM00486; POLBc; 1.  
 CC TIGRFS; TIGR00592; pol2; 1.  
 CC PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
 CC TRANSFERASE; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; 3D-structure.  
 KW SEQUENCE 775 AA; 89781 MW; 71525D63C0C9CD9D CRC64;  
 SQ  
 Query Match 79.6%; Score 3204.5; DB 1; Length 775;  
 Best Local Similarity 76.6%; Pred. No. 5.4e-167;  
 Matches 592; Conservative 89; Mismatches 89; Indels 3; Gaps 2;  
 QY 1 MIFDTDTKQKPIIRIFKKEGEFKIEDPHQPIYALLKDDSAIDKAIKGRHG 60  
 DB 1 MILDTDYITENGKPIRVIFKKEGEFKIEYDRTPEFYALLKDDSAIEDKVKYAKRHG 60  
 QY 61 KIVRVVDVAVKVKKFLGRDVEVWKLIFEPHQDVPALRGKIREHPAVIDIYEYDIPFAKRY 120  
 DB 61 TVVKVKAQKVKQKFLGRFIEVWKLIFVNHQDVPALRGKIREHPAVIDIYEYDIPFAKRY 120  
 QY 121 LIDKGLIPWEGDEELKMAFDIETPYHEGDFGKEIIMISYADEERARVITWKNIDLPY 180  
 DB 121 LIDKGLIPWEGDEELTMAFDIETLYHEGEFPGFPLMSIYADGSEARVITWKKIDLPY 180  
 QY 181 VDVVSNRERMIKRFVQIVREKDPDVLITVYNGDNFPLYLKRAEKLGVTLGLGRDKSHP 240  
 DB 181 VDVVSTEKEMIKRFRVREKDPDVLITVYNGDNFDFALVKRCBELGKIFTLGRDGS--E 238  
 QY 241 PKIHRMGOSFAVEIKGRIFHDFPVRRTINLPVTLVAVYAVLGKTKSLGAEETAAI 300  
 DB 239 PKIQMGDRFAVEVKGRIHFDLPVIRRTINLPVTLVAVYAVFGEKFKVYAEETAQA 298  
 QY 301 WTERESMKLAQYSEDARATYELGKEPFMEAEALIGOSVMDVSRSSGTGNVWYLL 360  
 DB 299 WESGEGLEKARVISEDARVTVELGREFPFMEAQLRSLIGSLNDVSRSSGTGNVWYLL 358  
 QY 361 RVAYERNELAPNKPDEEYRRRLRTTYLGGYKKEPERGLWENTYDPRCLYPSIIVTHN 420  
 DB 359 RKAYKNELAPNKPDERELARR-RGGYAGGYKKEPERGLWMDNIVYDLSRSLYPSIIVTHN 417  
 QY 421 VSPDPTLERGGCKNYDAPVGVGKFCDFPGFIPSLILGELITVRQIKKMKATIDPIEKK 480  
 DB 418 VSPDPTLRGGCKEYDAPVGVGKFCDFPGFIPSLILGDLLEERQIKKMKATVDPIEKK 477  
 QY 481 MLDYRQRAVKLANSYGYNGYVPKARWYKCEAESVTAWGRHYIEMTINKIEBEKFGFKVL 540  
 DB 478 LLDYRQRAIKLANSPYGYNGYVYAKARWYKCEAESVTAWGRYIEMWIRELEEKFGFKVL 537

QY 541 YADTDGFYATIPGKPEIKKKAKEFLKYINSKLPGLLELEYEGFYLRGFFVAKRYAVI 600  
 DB 538 YADTDGHLAIPGADATVKKKAKEFLKYINPKLPGLLELEYEGFYLRGFFVKKYAVI 597  
 QY 601 DEGRITTRGLEVVVRDWSIAKETAQKVLKEDSVKAEIVKDVVEEIAKYOVL 660  
 DB 598 DECKITTRGLEIVVRDWSIAKETAQKVLKEDSVKAEIVKDVVEEIAKYSKYEVP 657  
 QY 661 EKLVIHQITKDLSEYKAIQPHVAIAKLAAGKIKVRPGTIISYIVLRGSGKISDRVILL 720  
 DB 658 EKLVIHQITKDLKDYKATGPHVAIAKLAAGKIKVRPGTIISYIVLRGSGKISDRVILL 717  
 QY 721 SEYDPKHKHYPDYYIENQVLPVLRILEAPGYKEDLKYSQSSKQVGLDMLK 773  
 DB 718 DEFDPTKHRYDAEYIENQVLPVLRILEAPGYKEDLKYSQSSKQVGLDMLK 770  
 RESULT 5  
 DPOL PYRAB STANDARD; PRT; 771 AA.  
 ID AC P77916; P77932;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE DNA polymerase 1 (EC 2.7.7.7) (Pab polymerase).  
 GN POLI OR POL OR PYRAB17200 OR PAB1128.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]\_TaxID=29292;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ST 855;  
 RA Cambon M., Querellou J., Barbier G., Dietrich J., Forterre P.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GE23;  
 RA Cambon M., Querellou J., Bouyoub A., Raguene G., Barbier G.,  
 Forterre P., Dietrich J.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GE5 / Orsay;  
 RX PubMed=12622808;  
 RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
 Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,  
 Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
 RT "An integrated analysis of the genome of the hyperthermophilic  
 archaeon Pyrococcus abyssi".  
 RL Mol. Microbiol. 47:1495-1512(2003).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + (DNA)(N)  
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; Z54174; CAA90888.1; -  
 CC EMBL; Z54173; CAA90887.1; -  
 CC PIR; C75023; C75023.  
 CC HSP; P56689; ITGO.  
 CC InterPro; IPR006172; DNA\_pol\_B.  
 CC InterPro; IPR006134; DNA\_pol\_B\_dom.  
 CC InterPro; IPR006133; DNA\_pol\_B\_exo.  
 CC InterPro; IPR004578; Pol2.  
 CC Pfam; PF00136; DNA\_pol\_B\_1.

DR Pfam: PF03104; DNA\_pol\_B\_exo; 1.  
DR PRINTS; P00106; DNAPOLB.  
DR SMART; SM00486; POLB; 1.  
DR TIGRFAMS; TIGR00592; pol12; 1.  
DR PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
KW Transferase; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding; Complete proteome.  
FT VARIANT 532 532 R -> S (IN STRAIN GE23).  
FT VARIANT 533 554 HE -> PN (IN STRAIN GE23).  
FT VARIANT 756 756 K -> R (IN STRAINS GE23 AND ST 855).  
FT CONFLICT 263 263 V -> A (IN REF. 1).  
FT CONFLICT 277 277 A -> T (IN REF. 1).  
FT CONFLICT 281 281 F -> V (IN REF. 1).  
FT CONFLICT 320 320 F -> S (IN REF. 1).  
FT CONFLICT 339 339 Q -> H (IN REF. 1).  
FT CONFLICT 359 359 R -> T (IN REF. 1).  
FT CONFLICT 391 391 K -> N (IN REF. 1).  
SQ SEQUENCE 771 AA; 89496 MW; 110A87045A8A5522 CRC64;  
  
Query Match 78.6%; Score 3164.5; DB 1; Length 771;  
Best Local Similarity 76.1%; Pred. No 88-165;  
Matches 588; Conservative 89; Mismatches 93; Indels 3; Gaps 2;  
  
QY 1 MFTDTYITKQKPIIRIFKNGKGFKIELDPHPQYIYALLKDSADIEKAIGERHG 60  
DB 1 MIIDADYITDGGKPIIRIFKNGKGFKVEYDRTFRPIYIALLKDSADIEKITAERHG 60  
  
QY 61 KIVRVVDVAVKKELGRDVEYVWKLIFEPQDPVAPALRGKIREHPAVIDIYVDIPAKRY 120  
DB 61 KIVRITEVEKVKKFLGRPIEIVWKLXLEHPQDPVAPALRGKIREHPAVIDIYVDIPAKRY 120  
  
QY 121 LIDKGLIPMEGDEELKMAFDIETFYHGDGFGKGEIIMISYADEEEARVITWKNIDLPY 180  
DB 121 LIDKGLTPMEGNEELTFIYDITLYHEGEFGKGPIMISYADEEGAKVITWKSIDLPY 180  
  
QY 181 VDVVSNRERMIKRFQIVREKDPDVLITNGDNFPLPYLTKRAEKLGTWLLGRDKEHPE 240  
DB 181 VEVVSEREMIRLVKIVREKDPDVLITNGDNFPLPYLTKRAEKLGTWLLGRD--NSE 238  
  
QY 241 PKIHRMGDSFAVEIKRIFHDFLPVVRRTINLPTVTLAEVAVLGTGKSKLGADEIAAI 300  
DB 239 PKQCRMGDSLAVELKIRIFHDFLPVVRRTINLPTVTLAEVAVLGTGKSKVYAHIAEA 298  
  
QY 301 WETESMKLAQYSMEDATATVETGKEPPMAEALKIGQSVWDSRSSGNLVEWILL 360  
DB 299 WETGKGLERVAKYSMEDAKVTTELKGFPPMAEALRLVGPVWDSRSSGNLVEWFL 358  
  
QY 361 RVAYERNELAPKPDDEEYRRRLRTTYLGGYVKEPERGLWENITVLDPRCLYPSIIVTHN 420  
DB 359 RXAYERNELAPKPDDEERYERLRRESYEGYVKEPKGLWEGIVSLDFPSLYPSIIVTHN 418  
  
QY 421 VSPDTLREGCKYDVAPIVGKPKDPGPTPSILGELITNWRQIKKKKATIDPIEKK 480  
DB 419 VSPDTLNRNCKEYDVAPQVGHFRFCDFGPTPSLLGNLLEERQIKKRMKESKDPVEKK 478  
  
QY 481 MLDYRORAVKLHANSYGYMGYPKARWYSKECAESVTAGRHVYIEMTIKEEKGFKVL 540  
DB 479 LLDYRORAKILANSYGYGYGAKARWYSKECAESVTAGRGYIDLVRRELESR-GFKVL 537  
  
QY 541 YADTGFYATIGEPETIKKAKFLKYNSKLPGLLEVEGYFLRGFFVAKRYAVI 600  
DB 538 YIDTGLYATIPGAKHEIEKALKAFVEYNSKLPGLLEVEGYFGYRGFFVTKKLYALI 597  
  
QY 601 DEEGRTTTRGLVWRDWESEIAKETQAKVLBAIKEDSVKAVEVWQVDEIAKQVPL 660  
DB 598 DEEGKIVTGLVWRDWESEIAKETQAKVLBAIKGNVDEAVKLVKEVTEKLSKYEIIP 657  
  
QY 661 EKLVIHQITKOLSEYKATGPHVAIAKLAAGKIKVRPTTIISYIVLRGSKISDRVILL 720  
DB 658 EKLVIYEQITRPLSEYKATGPHVAIAKLAAGKVKVPGWIGYIVLRGSDPISKRAIAI 717  
  
QY 721 SEYDPKKHKYDPDYIENOVLPVLRILLEAFGYRKEDLKQSSKQVGLDAMLK 773

DB 718 EEPDPKKHKYDAEYIENOVLPVLRILLEAFGYRKEDLKQTKQVGLGAWLK 770  
  
RESULT 6  
ID DPOL\_PYRPU STANDARD; PRT; 775 AA.  
AC P80061; P95584;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA polymerase (EC 2.7.7.7) (Pfu polymerase).  
GN POL OR PF0212.  
OS Pyrococcus furiosus, and  
OS Pyrococcus woesei.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261, 2262;  
RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 1-12.  
RP SPECIES=P.furiosus; STRAIN=Vci / DSM 3638 / ATCC 43587 / JCM 8422;  
RA MEDLINE=93181200; PubMed=8441634;  
RA Uemori T., Ishino Y., Toh H., Asada F., Kato I.;  
RT "Organization and nucleotide sequence of the DNA polymerase gene from  
RL the archaean Pyrococcus furiosus.";  
RN Nucleic Acids Res. 21:259-265(1993).  
[2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.furiosus; STRAIN=Vci / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE OF 396-418 FROM N.A.  
RC SPECIES=P.furiosus;  
RA MEDLINE=92107689; PubMed=1762925;  
RA Mathur E.J., Adams M.W.W., Callen W.N., Cline J.M.;  
RT "The DNA polymerase gene from the hyperthermophilic marine  
RL archaeobacterium, Pyrococcus furiosus, shows sequence homology with  
RT alpha-like DNA polymerases.";  
RN Nucleic Acids Res. 19:6952-6952(1991).  
[4]  
RP SEQUENCE OF 396-502 FROM N.A.  
RC SPECIES=P.furiosus;  
RA Mathur E.J., Adams M.W.W., Callen W.N., Cline J.M.;  
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P.woesei; STRAIN=DSM 3773;  
RA MEDLINE=98432910; PubMed=9758761;  
RA Dabrowski S., Kur J.;  
RT "Cloning and expression in Escherichia coli of the recombinant his-  
RL tagged DNA polymerases from Pyrococcus furiosus and Pyrococcus  
RL woesei.";  
RN Protein Expr. Purif. 14:131-138(1998).  
[6]  
RP SIMILARITY TO OTHER POLYMERASES.  
RA MEDLINE=92253396; PubMed=1579479;  
RA Forterre P.;  
RT "The DNA polymerase from the archaeobacterium Pyrococcus furiosus does  
RL not testify for a specific relationship between archaeobacteria and  
RL eukaryotes.";  
RN Nucleic Acids Res. 20:1811-1811(1992).  
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
CC EXHIBITS 3, TO 5' EXONUCLEASE ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA} [N].  
CC -1- SUBUNIT: Monomer.  
CC -1- BIOTECHNOLOGY: Because pfu DNA polymerase exhibits the lowest  
CC error rate of any thermostable DNA polymerase studied, it is  
CC routinely used for PCR. It is sold by Promega.  
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration





DR PRINTS; PRO0106; DNAPOLB.  
DR PRINTS; PRO0379; INTEIN.  
DR SMART; SM00305; HintC; 1.  
DR SMART; SM00306; HintN; 1.  
DR SMART; SM00486; POLBc; 1.  
DR TIGRFAMS; TIGR01443; intein\_Cterm; 1.  
DR TIGRFAMS; TIGR01445; intein\_Nterm; 1.  
DR TIGRFAMS; TIGR00592; Pol2; 1.  
DR PROSITE; PS00116; DNA POLYMERASE B; 1.  
DR PROSITE; PS00818; INTEIN C TER; 1.  
DR PROSITE; PS00819; INTEIN ENDONUCLEASE; 1.  
DR PROSITE; PS00817; INTEIN N TER; 1.  
KW Transferase; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;  
KW Protein splicing; Intrin homing.  
FT CHAIN 1 492 DNA POLYMERASE, 1ST PART.  
FT CHAIN 493 1029 ENDONUCLEASE PI-PSPI.  
FT CHAIN 1030 1312 DNA POLYMERASE, 2ND PART.  
SQ SEQUENCE 1312 AA; 152852 MW; 862518805641D26A CRC64;  
  
Query Match 72.9%; Score 2934.5; DB 1; Length 1312;  
Best Local Similarity 45.6%; Pred. No. 58-152; Mismatches 87; Indels 539; Gaps 3;  
Matches 597; Conservative 86;  
  
QY 1 MIFDTYITKDGKPIIRIPKENGSEFKIELDPHFQPIYVALLKDDSAIDEIKAIKGERHG 60  
DB 1 MILDADYITDCKPIIRIPKENGSEFKVEYDRNFRPIYVALLKDDSQIDEVRKITAERHG 60  
  
QY 61 KIVRVVDVAVKKKPLGRDVEVWKLIFEPQDPVPAIRKIRHPAVIDIYEDIPFAKRY 120  
DB 61 KIVRIIDAEKVRKPLGRDIEVWRLYFEHPQDPVPAIRKIRSHSAVIDIFEDIPFAKRY 120  
  
QY 121 LIDKGLIMPGEDEELKMAFDIETVHEGDEFGKEIIMISVADDEEARVITWKNIDLPY 180  
DB 121 LIDKGLIMPGEDEELKMAFDIETVHEGDEEFAKGIIMISVADDEEARVITWKNIDLPY 180  
  
QY 181 VDVNSNERMIRFVQIVREKDPDLVITNGDNFDPYLIKRAEKLGTWLLGRDKHEPE 240  
DB 181 VEWVSEREMIRFLKIVREKDPDVIITNGDSFDPYLVKRAEKLGTPLGRDGS--E 238  
  
QY 241 PKIHRMGDSFAVEIKGRHFDLPVVRRTINPTVTLAVVAVLGKTKSLGABEIAAI 300  
DB 239 PKMQLGDMTAVEIKGRHFDLYHVRRTINPTVTLAVVAVLGKTKSLGABEIAEA 298  
  
QY 301 WTEESMKLAQYSMEDAPATVELAKGEPFPMABELAKIGQSVDPVSRSTGNLVWYLL 360  
DB 299 WETGKGLERVAKYSMEDAKVTVELGREFPMEQSLVQQLWDVSRSTGNLVWYLL 358  
  
QY 361 RVAYERNELAPKPDDEEYRRRLRTYLGYYVKEBERGLWENITVLDPRCLYPSIIVTHN 420  
DB 359 RXAYERNELAPKPDDEEYRRRLRTYLGYYVKEBERGLWGLVSLDPRSLYPSIIVTHN 418  
  
QY 421 VSPDTLREGCKNYDVAPITVGYKFCDFPGFIPISILGELITWQRIKKMKATIDPIEKK 480  
DB 419 VSPDTLNREGCREYDVAPVGVGHKFCDFPGFIPISILKRLDLDERQIKKMKASKDPIEKK 478  
  
QY 481 MLDYRQAVK----- 490  
DB 479 MLDYRQRAIKILANSILPEWVPLIKNGKVIKIFRIGDFVDGLMKANQKVKTKGTDEYLE 538  
  
QY 491 ----LHA----- 493  
DB 539 VAGIHAFSFRDSSKKARVMAKAVIRHRYSGNVRIVLNSGRKIITTEGHSIFVYRNGDL 598  
  
QY 494 ----- 493  
DB 599 VEATGEDVKIGDLAVPRSVNLPEKRLNIVELLNLSPEETDIILTIPIVKGKRNPFK 658  
  
QY 494 ----- 493  
DB 659 GMLRTLRFGBEKRVRTASVRLHLENLYRLKIGVDIIDKGLKRYTLYEKIVDV 718  
  
QY 494 ----- 493

Db 719 VRYNGKREYLVEFNARVDVISLMBEELKEWRIGTRNGFRMGTFVDIDEDFAKLLGYV 778  
QY 494 ----- 493  
Db 779 SEGSAKKWKQTCGWSYTVRLYNENDEVLDDEHLAKKFFGKVRGKNVVEIPKWAYII 838  
QY 494 ----- 493  
Db 839 FESLCGTLAENKRVPEVIFTSSKGVRAWFLEGYFIGDGVHPKSRVRLSTKSELLVGLV 898  
QY 494 ----- 493  
Db 899 LLNLSGVSAIKLGYDSGVYRVVNEELKFTYRKKKNVHSHVPOKILKETFQKVKFQK 958  
QY 494 ----- 493  
Db 959 NISYKKPRELVENGKLDREKAKRIEWLLNGDIVLDRAWVEIKREYYDGYVDLSVDENF 1018  
QY 494 -----NSYGYMGYPKAWYSKECAESVTANGRYHVIETIKIEIEKCFKVLVAD 543  
Db 1019 LAGGFYAHNSYGYGYAKARWYCKECAESVTANGRYHVIETIEFVRKELBEKFGFKVLIYD 1078  
QY 544 TDGFYATIPGKEPETIKKAKFELKYNKLPGLLELEYEGFYDLRGFFVAKKRYAVIDEE 603  
Db 1079 TDGLYATIPGKPEEIKKALEFVDYINAKLPGLLELEYEGFYVVGFFVTKKYYALIDEE 1138  
QY 604 GRITTRGLEVVRDWSIEIAKETOAKVLEALLKEDSVKAEIVKDVVEELAKYQVLEKL 663  
Db 1139 GKITTRGLEVVRDWSIEIAKETOAKVLEALLKHNVEEAVKIVKTEKLSKYEIPPEKL 1198  
QY 664 VIHEQITKDISYKAIGPHVAIAKRAKIGKIVRPGTIISYIVLRSGKISDRVILLSEY 723  
Db 1199 VIVEQITRPLHEYKAIGPHVAVAKLAARGVKVPCGWIGYIVLRGDPISKEAILAEF 1258  
QY 724 DPKKHKYDPPYYIENOVLPVAILFAFGYRKEDLYKSSKQVGLDAWL 772  
Db 1259 DLKHKHYDAEYYIENOVLPVAILFAFGYRKEDLFWQTKOTGLTAWL 1307  
  
RESULT 8  
DPOL\_PYRHO  
ID \_DPOL\_PYRHO STANDARD; PRT; 1235 AA.  
AC OS9610;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA polymerase (EC 2.7.7.7) [Contains: Phc pol intein (Phc Pol I  
intein)]  
GN POL OR PH1947 OR PHB047.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=96344137; PubMed=9679194;  
RA Kawarayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
Funahashi T., Tanaka T., Kudo H. Y., Yamazaki J., Kishida N., Oguchi A.,  
Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
+ {DNA}(N).  
CC -!- PIM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)  
FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP000007; BAA31074.1;
CC PIR; C71210; C71210.
CC HSP; P56689; 1TGO.
CC InterPro; IPR006172; DNA_pol_B.
CC InterPro; IPR006134; DNA_pol_B_dom.
CC InterPro; IPR006133; DNA_pol_B_exo.
CC InterPro; IPR003586; Hedgehog_HintC.
CC InterPro; IPR003587; Hedgehog_HintN.
CC InterPro; IPR006141; Intein.
CC InterPro; IPR006142; Intein.
CC InterPro; IPR004042; Intein_endonuc.
CC InterPro; IPR004578; Pol2.
CC Pfam; PF00136; DNA_pol_B; 1.
CC Pfam; PF03104; DNA_pol_B_exo; 1.
CC PRINTS; PR00106; DNAPOB.
CC PRINTS; PR00379; INTEIN.
CC SMART; SM00305; HintC; 1.
CC SMART; SM00306; HintN; 1.
CC SMART; SM00486; POLBc; 1.
CC TIGRFAMs; TIGR01443; intein_Cterm; 1.
CC TIGRFAMs; TIGR01445; intein_Nterm; 1.
CC TIGRFAMs; TIGR00592; pol2; 1.
CC PROSITE; PS00116; DNA_POLYMERASE_B; 1.
CC PROSITE; PS00818; INTEIN_CTER; 1.
CC PROSITE; PS00819; INTEIN_ENDONUCLEASE; 1.
CC PROSITE; PS00817; INTEIN_NTER; 1.
CC Transferrase; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Nucleolytic cleavage; Protein splicing;
CC Complete proteome.
CC CHAIN 1 492 DNA POLYMERASE, 1ST PART (POTENTIAL).
CC FT CHAIN 493 PHO POL INTEIN (POTENTIAL).
CC FT CHAIN 953 DNA POLYMERASE, 2ND PART (POTENTIAL).
CC SQ SEQUENCE 1235 AA; 143086 MW; 730C7AA14873CCE4 CRC64;

Query Match 71.7%; Score 2886; DB 1; Length 1235;
Best Local Similarity 46.5%; Pred. No. 2e-149;
Matches 574; Conservative 107; Mismatches 89; Indels 464; Gaps 5;

QY 1 MIFDTYITKDGPIIRIFPKENGSEKIFELDPHPQPIYVALLKDSADDEIKAKGERHG 60
Db 1 MILDADYITEDGPIIRIFPKENGSEKIFELDPHPQPIYVALLKDSADDEIKAKGERHG 60
QY 61 KIVRVVDVAVKXKFLGRDVEVWKLIFEPHPQDPVPAIRKIRHPAVIDIYEDIPFAKRY 120
Db 61 KVRVIVETKIQKFLGRPIEVWKLIVLEHPQDPVPAIRKIRHPAVIDIYEDIPFAKRY 120
QY 121 LIDKGLIPMEGDEBELKMAFDIETPYHGEDEFGKGEIIMISYADDEEAAVITWKNIDLPY 180
Db 121 LIDKGLTPMEGNEKLTFLAVDIETLYHEGEFEFGKGPVIMISYADDEEAGKVITWKNIDLPY 180
QY 181 VDVVSVNERMIKFPVQIVAREKDPDLITVNGDNFDLPYLTKAEKLGVTLLIGRKERPE 240
Db 181 VEVVSVNERMIKFLIRVIREKDPDLITVNGDNFDLPYLTKAEKLGVTLLIGRKERPE 240
QY 241 PKIHRGDSFAVIEKGRHIFDLFPVVRRTINLTPTYLEAVVEAVLTKTSKAGAEIAAI 300
Db 239 PKQKMGDSLAVIEKGRHIFDLFPVVRRTINLTPTYLEAVVEAVLTKTSKAGAEIAAI 298
QY 301 WETESMKLAQVSMEDARATYELGKFFPMEALAKLGOSVWDYSRSTGNLVEWYLL 360
Db 299 WETGEGLERVAKVSMEDAKTYELGREFFPMEALQARLVQFPVWDVSRSTGNLVEWYLL 358
QY 361 RVAVERNELAPNKPDEEYERRRLRTTYLGGYVKEPERGLWENTTYLDFRCLYPSIIVTN 420
```

```
Db 359 RKAVERNELAPNKPDEEYERRRLRESYEGGYVKEPKLWEGIVSLDFRSLYSIIITN 418
QY 421 VSPDTLREGCKNDVAPIVGYKFCDFPGFIPSLIGELITMRQETKQKWKATIDPIEKK 480
Db 419 VSPDTLREGCEEYDVAPKVGHRFCDFPGFIPSLIGQLLEERQKIKRMRKESKDPVEKK 478
QY 481 MLDYRQRAVKL----- 491
Db 479 LLDYRQRAIKLANSILPDEWLPIVENEKRVFVKIGDFIDRETEENAEVRKRDGETELLE 538
QY 492 ----- 491
Db 539 VKDLKALSFNRETCKSELKKVYKALIRHRYSGKVYSIKLKGRRIKITSGHSLFSVQKGL 598
QY 492 ----- 491
Db 599 VKVRGDELKPGDLVVPGRKLPEKQVNLVELLLKLPEBETSNIIVMIPVKGRKNFFK 658
QY 492 ----- 491
Db 659 GMLKTLVYFGEGERPRTAGRYLXHLERLGYVVKLRRCCEVLDWESLKRKYKLYETLIKN 718
QY 492 ----- 491
Db 719 LKYNCSRAYVVFENSLRDVSLMPIELKEWIIICEPRGPKIGTIFDVDDSFALGLYII 778
QY 492 ----- 493
Db 779 SSGDVEKRVKPHSKDQNVLEDAKLAELFKGVRRGRGYIEVSGKISHAIFRVLAEGR 838
QY 494 ----- 493
Db 839 IPEFTFTSPMDIKVAPLKLNGNABELAFSTKSELLVNLILLNSIGVSDIKIEHEKV 898
QY 494 ----- 500
Db 899 YRVYINKESSNGDIVLDSVESIEVEKGYVYVDSVEDNENFLVGLGLYAHNSYGY 958
QY 501 GYPMARWYKECAESVTAMGRHYIEMTKIEBKFGKVLVADTDFGFTATITCEKP-EPI 559
Db 959 GYKASWYKECAESVTAMGRQYIDLVRELEAR-GFKVLYITDGLYATIPGVKDWEEV 1017
QY 560 KKKAKFLKYNKSLPGLLEVEGFLRGFPFAKRVAVIDEBGRITTRGLEVVRDWS 619
Db 1018 KRSALEFVYINSLKGLVLEVEGFLRGFPFAKRVAVIDEBGRITTRGLEVVRDWS 1077
QY 620 EIAKETQAKVLEAILKEDSVKAVEIVKDVVEIAKYQVPLEKLVHIEQITKDLSEYKAI 679
Db 1078 EIAKETQARVLEAILKHGNVBEAVKVKDTEKLTNYEVPPEKLVIEQITRPINEYKAI 1137
QY 680 GPHVATAKRLAAGIKVRPGTIIISYVLRSGKISDRVILLSEYDPKHKYDPPVYIENQ 739
Db 1138 GPHVAVAKRLMARGIKVKGPMVIGYIVLRGDPISKRAISIEFDFPRKHKYDAEYIENQ 1197
QY 740 VLPVAVRIIEAFGRYRKEDLKYSSKQVGLDANLK 773
Db 1198 VLPVAVRIIEAFGRYRKEDLKYSSKQVGLDANLK 1231
RESULT 9
ID DPOL THEFM STANDARD; PRT; 1523 AA.
AC P74918;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) (pol Tfu) [Contains: Endonuclease PI-TfuI
DE (EC 3.1.-.-) (Tfu pol-1 intein); Endonuclease PI-TfuII (EC 3.1.-.-)
DE (Tfu pol-2 intein)].
GN POL.
OS Thermococcus funiculans.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
```

OX NCBI\_TaxID=46540;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ST557;  
RA Sambon M., Querellou J.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RV [2]  
RV CHARACTERIZATION OF INTENS.  
RP STRAIN=ST557;  
RX MEDLINE=20112788; PubMed=10644683;  
RA Saves I., Ozanne V., Dietrich J., Maason J.-M.;  
RT "Intens of Thermococcus fumicolans DNA polymerase are endonucleases  
RT with distinct enzymatic behaviors";  
RL J. Biol. Chem. 275:2335-2341 (2000).  
CC -!- FUNCTION: PI-TfuII recognizes and cleaves a minimal sequence of 16  
CC base pairs (bp) on supercoiled DNA with either Mn(2+) or Mg(2+) as  
CC cofactor. It cleaves linear DNA only with Mn(2+) and requires a  
CC 19-bp minimal recognition sequence. The optimal temperature for  
CC activity is 70 degrees Celsius.  
CC -!- FUNCTION: PI-TfuII is a highly active homing endonuclease using  
CC Mg(2+) as cofactor. Its minimal recognition and cleavage site is  
CC 21 bp long either on linear or circular DNA substrates. Its  
CC endonuclease activity is strongly inhibited by the 3' digestion  
CC product, which remains bound to the enzyme after the cleavage  
CC reaction. The optimal temperature for activity is 70 degrees  
CC Celsius.  
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + [DNA] (N).  
CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
CC A POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION  
CC (INTENS) FOLLOWED BY PEPTIDE LIGATION.  
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
CC -!- SIMILARITY: IN THE INTEN SECTION; BELONGS TO THE HOMING  
CC ENDONUCLEASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z69882; CAA93738.1; -  
CC HSP; P56689; ITGO.  
CC REBASE; 4500; PI-TfuII.  
CC REBASE; 4501; PI-TfuII.  
CC InterPro; IPR006172; DNA pol B.  
CC InterPro; IPR006134; DNA pol\_B\_dom.  
CC InterPro; IPR006133; DNA pol\_B\_exo.  
CC InterPro; IPR003586; Hedgehog\_hintC.  
CC InterPro; IPR003587; Hedgehog\_hintN.  
CC InterPro; IPR006141; Intein.  
CC InterPro; IPR006142; INTEIN.  
CC InterPro; IPR004042; Intein\_endonuc.  
CC Pfam; PF00136; DNA\_pol\_B; 2.  
CC Pfam; PF03104; DNA\_pol\_B\_exo; 1.  
CC PRINTS; P00379; INTEIN.  
CC SMART; SM00305; HintC; 2.  
CC SMART; SM00306; HintN; 2.  
CC SMART; SM00486; POLB; 1.  
CC TIGRfams; TIGR01443; intein\_Cterm; 2.  
CC TIGRfams; TIGR01445; intein\_Nterm; 2.  
CC TIGRfams; TIGR0592; pol2; 1.  
CC PROSITE; PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.  
CC PROSITE; PS00818; INTEIN\_C\_TER; 2.  
CC PROSITE; PS00819; INTEIN\_ENDONUCLEASE; 2.  
CC PROSITE; PS00817; INTEIN\_N\_TER; 2.  
CC Transferrase; DNA-directed DNA polymerase; DNA replication;  
CC DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;  
CC Protein splicing; Intronic homing; Magnesium; Manganese.  
CC CHAIN 1 406 DNA POLYMERASE, 1ST PART.

FT	CHAIN	407	766	ENDONUCLEASE PI-TFUII.
FT	CHAIN	767	900	DNA POLYMERASE 2ND PART.
FT	CHAIN	901	1282	ENDONUCLEASE PI-TFUII.
FT	CHAIN	1283	1523	DNA POLYMERASE, 3RD PART.
SO	SEQUENCE	1523	AA; 175917	MW; 7A2AC236BF2E5F5 CRC64;
Query Match				
Best Local Similarity 39.0%; Pred. No. 4.1e-145;				
Matches 593; Conservative 86; Mismatches 91; Indels 752; Gaps 4;				
QY	1	MIFDTYITKDGKPIIRIKFNGEKPILDPHFQPIYIYALLKDDSAIDEIKAKGERHG	60	
DB	1	MILDTYITTEDGRPIRVFKENGEPKIEYDRDFEYIYALLKDDSAIDEIKAKTASRHG	60	
QY	61	KIVRVVDVAVKVKKFLGRDVEVWKLIFFHPQVPALRGKIRHPAVIDIYVDIPFAKY	120	
DB	61	TTVRVVRAGKVKKFLGRPIEVWKLYFTHPQVPALRKIRHPAVIDIYVDIPFAKY	120	
QY	121	LIDKGLIPMEGBELKMAPDIETFYHGEDEPGKEIIMISYADBEAEAVITWKNIDLPY	180	
DB	121	LIDKGLIPMEGBELKMLAFDIETLYHGEDEPAEGPILMISYADBEAEAVITWKNIDLPY	180	
QY	181	VDVVSNEREMIKFVOIVREKDPDVLITVNGDNFDLPVLIKRAEKLGVTLGGRDKEHPE	240	
DB	181	VDVVSNEKEMIKFVKVKEKDPDLITVNGDNFDLPVLIKRAEKLGVTLGGRDKEHPE	240	
QY	241	PKIHRMGDSFAVEIKRIHFDLPFPVVRITINLPTVTEAVYPAVLGKTSKLGABEIAAI	300	
DB	239	PKIQRMGDRFAVEIKRIHFDLPVIRITINLPTVTEAVYPAVLGKTSKLGABEIAAI	298	
QY	301	WETEESEKLAQVSMEDARATVELGKEFPFMAELAKLIGQSVWDSRSTGNLVEWYLL	360	
DB	299	WETGEGLERVARYSMEDAKVTVELGKEFPFMAELAKLIGQSVWDSRSTGNLVEWYLL	358	
QY	361	RVAYERNELAPNKPDEEYRRLRTYLGYYKVEPERGLWENITVLDPRC	410	
DB	359	RKAVERNELAPNKPDSRELERR-RGGYAGGYKVEPERGLWENIAYLDFRCHPADTKVIK	417	
QY	411		410	
DB	418	KGQVNVISEVREGDYVLGIDGKQVQRVWEYDEGELVNVINGLKTNPKNLPPVRRTERQ	477	
QY	411		410	
DB	478	TAIRDSLAKSFLTKVKVGLITLPLFEKIGKIGTEREDVPEEILKGLAGIILAEGLTLRK	537	
QY	411		410	
DB	538	DVEYFDSSRGKRVSHQVNRVEITVGAQBEDFORRIVYIPERLFVTPSVYRKKNATITF	597	
QY	411		410	
DB	598	KVAKKEVYLVRREIMDGIENLHAPSLRGFFBGDSGVNKRVTVVVNOGTNNEMKIEVVS	657	
QY	411		410	
DB	658	KLLNKLGIPIHRYTYDYTEREKTMTTHILEIAGRDGLIFQITGVFISTEKNMALEAIR	717	
QY	411		410	
DB	718	NREVNRLNNAFYTLADFTAKTEYKGVYDLTLEGTYPYFANGILTNLSYPSIIISHN	777	
QY	421	VSPDTLEREGCKNVDVAPIGVYKFCDFPGFIPSLBELITNRQBIKKKMKATIDPIEKK	480	
DB	778	VSPDTLNRGCGEYDEAPQVGHRCDFPGFIPSLGDLDERQKVKKMKATIDPIEKK	837	
QY	481	MLDYQRAVKLHANSYGYGMGYPKARWYKCECAEAVTAWGRHYIEMTKEIEEKGFKVL	540	
DB	838	LIDYQRAIKILANSFYGYGYKARWYKCECAEAVTAWGRHYIEMTKEIEEKGFKVL	897	
QY	541	YAD	543	
DB	898	YADSVTGDTEVTRIRNGRIEFVPIEKLPERVDHVRGEKEYCVLGGVEALTLDNRGLVWK	957	

```

QY 544 ----- 543
Db 956 KVEYVNRHTDKRIYRVWFTNSWYLDVTDHSLGLYNTSKVPGKPLKRLVEVRPEEL 1017
QY 544 ----- 543
Db 1018 GGVKSLITNRPARTIKANPIAVKLMELIGLVGDNWGGQSNWAKYVYVGLSCGLDKA 1077
QY 544 ----- 543
Db 1078 EIERKVLNPLREASVISNYDKSKGDVSIILSKMLAGFWKVFYFDENGKKAIPSFMPNLP 1137
QY 544 ----- 543
Db 1138 REYIEAFLRGLFSADCTVSLRRGIPEIRLTYSNRELSDAVRKLLWLVGVNSLFTETKPN 1197
QY 544 ----- 543
Db 1198 RYLEKSGTHSIHVRINKHRFADRLGFLIDKSTKLSNGLGHTNKKRAYKYDFDLVYP 1257
QY 544 ----- 571
Db 1258 RKIEETTYGYVYDIEVEGTHRFANGILVHNTDGFATIPGADAETVKKAREFLNYIN 1317
QY 572 SKLPGLLELEYEGFYLRGFFVAKRYAVIDEGRITTRGLEVVRDSEIAKETOAKVLE 631
Db 1318 PKLPGLLELEYEGFYRGGFVTKKYAVIDEGRITTRGLEVVRDSEIAKETOAKVLE 1377
QY 632 AILKEDSVKAEIVKQVVEEIAKQVPLEKLVHEQITKOLSEYKALGPHVAIAKRLAA 691
Db 1378 AILRHGDVSEARIVKEVTEKLSKVEVPEKLVHEQITRSLKDYKAGPHVAIAKRLAA 1437
QY 692 KGIKVRPGTHIISVILRGSGKISDRVILLSEYDPPKKHYDDYVYENQVLPVAVLRILEAF 751
Db 1438 RGIVAPGTVISVILKSGRIGDRITPDEBDPTKHYDAEYVYENQVLPVAVLRILEAF 1497
QY 752 GYKEDOLKYOSSKQVGLGAWLK 773
Db 1498 GYKEDLYQKIQVGLGAWLK 1519

RESULT 10
DPOL PYRKO
ID DPOL PYRKO STANDARD; PRT; 1671 AA.
AC P7933;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) [Contains: Endonuclease P1-PkoI
DE (EC 3.1.1.-) (Pko pol-1 intein) (IVS-A); Endonuclease P1-PkoII
DE (EC 3.1.1.-) (Pko pol-2 intein) (IVS-B)].
GN POL.
OS Pyrococcus kodakaraensis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=69014;
RN [1]
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=KOD1;
RX MEDLINE=98027387; PubMed=9361436;
RA Takagi M., Nishioka M., Kakiwara H., Kitabayashi M., Inoue H.,
RA Kawakami B., Oka M., Imanaka T.;
RT "Characterization of DNA polymerase from Pyrococcus sp. strain KOD1
RT and its application to PCR."
PL Appl. Environ. Microbiol. 63:4504-4510 (1997).
RN [2]
CHARACTERIZATION OF INTEINS.
RC STRAIN=KOD1;
RX MEDLINE=98416198; PubMed=9742242;
RA Nishioka M., Fujiwara S., Takagi M., Imanaka T.;
RT "Characterization of two intein homing endonucleases encoded in the
RT DNA polymerase gene of Pyrococcus kodakaraensis strain KOD1."

```

```

RL Nucleic Acids Res. 26:4409-4412 (1998).
CC -!- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE
CC INTEIN MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY
CC ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE
CC INTEIN. P1-PKOI RECOGNIZES 5'-GATTAGTACCTGTACC-3' AND P1-PKOII
CC RECOGNIZES 5'-CAGTACTAGGTAC-3'. BOTH ARE THERMOSTABLE.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- PM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC -!- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
CC ENDONUCLEASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D29671; BAA06142.2; -.
CC PDB; 1GCX; 28-JAN-03.
CC REBASE; 3792; P1-PKOI.
CC REBASE; 3793; P1-PKOII.
CC InterPro; IPR006172; DNA_pol_B.
CC InterPro; IPR006134; DNA_pol_B_dom.
CC InterPro; IPR006133; DNA_pol_B_exo.
CC InterPro; IPR003586; Hedgehog_hintc.
CC InterPro; IPR003587; Hedgehog_hintN.
CC InterPro; IPR006141; Intein.
CC InterPro; IPR006142; INTEIN.
CC InterPro; IPR004042; Intein_endonuc.
CC InterPro; IPR004578; Pol2.
CC Pfam; PF00136; DNA_pol_B_3.
CC Pfam; PF03104; DNA_pol_B_exo; 1.
CC PRINTS; PR00379; INTEIN.
CC SMART; SM00305; HintC; 2.
CC SMART; SM00306; HintN; 2.
CC SMART; SM00486; POLBc; 1.
CC TIGRFAMs; TIGR01443; intein_Cterm; 2.
CC TIGRFAMs; TIGR01445; intein_Nterm; 2.
CC TIGRFAMs; TIGR00592; pol2_2.
CC PROSITE; PS00115; DNA_POLYMERASE_B; 1.
CC PROSITE; PSS0818; INTEIN_C_TER; 2.
CC PROSITE; PSS0819; INTEIN_N_TER; 2.
CC PROSITE; PS50817; INTEIN_N_TER; 2.
CC TRANSFERASE; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;
CC Protein splicing; Intron homing; 3D-structure.
CC CHAIN 1 406 DNA POLYMERASE, 1ST PART.
CC CHAIN 407 766 ENDONUCLEASE P1-PKO I.
CC CHAIN 767 851 DNA POLYMERASE, 2ND PART.
CC CHAIN 852 1388 ENDONUCLEASE P1-PKO II.
CC CHAIN 1389 1671 DNA POLYMERASE, 3RD PART.
CC SEQUENCE 1671 AA; 193490 MW; 4A17F1C9120EE455 CRC64;

Query Match 69.1%; Score 2782; DB 1; Length 1671;
Best Local Similarity 36.0%; Pred. No. 1.3e-143;
Matches 602; Conservative 82; Mismatches 86; Indels 900; Gaps 4;

QY 1 MIFDTDTITDGGKPIIRIFKKEGFKILDPHFQPIYVALLKDDSAIDEIKAKGERHG 60
Db 1 MILDITITDGGKPIIRIFKKEGFKILDPHFQPIYVALLKDDSAIEEVKKTIAERHG 60
QY 61 KIVRVDAVKKKFLGRDVEVWKLIFEPQDVPAALRGKIREHPAVIYIYEDIPFARY 120
Db 61 TWTVTKRVKKKFLGRDVEVWKLIFEPQDVPAALRGKIREHPAVIYIYEDIPFARY 120
QY 121 LIDKGLIPMGDDDELKMAFDITTFVHEGDFGKGIIMISYADEEAEARVITWKNIDLPY 180

```

Db 121 LIDKGLVPMGEDELKMLAFDIETLYHEGEBAEFGPILMSIYADBEGARVITMKNDLPY 180  
QY 181 VDVVSNREMIKRFVQIVREKDPDVLITVNGDNFDPYLIKRAKLGVTLLGRKEHPE 240  
Db 181 VDVVSTERMIKRFURVREKDPDVLITVNGDNFDPYLIKRAKLGVTLLGRKEHPE 238  
QY 241 PKIHRMGDSFAVEIKRIRHFDLPVVRRTINLPTTLEAVYEAVALGKTKSLGAEIAAI 300  
Db 239 PKIQMGDRFAVEIKRIRHFDLPVVRRTINLPTTLEAVYEAVALGKTKSLGAEIAAI 298  
QY 301 WETESMKLAQYSMEDARATVELGKFFPMABEAKLIGQSVWDVSSSTGNLVEWYLL 360  
Db 299 WETGENLVARYSMEDAKVTYELGKFFPMABEAKLIGQSVWDVSSSTGNLVEWYLL 358  
QY 361 RVAYERNELAPKPDDEEYERRLRTYLGYYVKEPERGLWENITYLDFRC 410  
Db 359 RVAYERNELAPKPDDEEYERRLRTYLGYYVKEPERGLWENITYLDFRC 417  
QY 411 410  
Db 418 KGKIINISEVQEGYVLGIDGQVRVKWWEYDYKGLVNLGLKCTPNHKLPPVTKBRQ 477  
QY 411 410  
Db 478 TRIRDSLAKSLTKVKKIIITPLPYEIGRATSENIPEEVLKGLAGILLASGTLEK 537  
QY 411 410  
Db 538 DVEYFDSRRKRRIHQYRVEITIGKDEEPRDRITYIPERLFGITPISSEKKTNAVTL 597  
QY 411 410  
Db 598 KVAKKNVYLKVEIMDNIBSLHAPSVLRGFFGDSVNRVRSIVATQGTNEWKILVS 657  
QY 411 410  
Db 658 KLLSQIGIPHQTYTYQYQENGKDRGRYILEITGKDLILFQTLGIFGERKNALLKALS 717  
QY 411 410  
Db 718 QREMANLENGYRVRSEFNVSTYEGVKYDUTLEGTPYFANGILTTHNSLPSIIITHN 777  
QY 421 VSPDTLREGCKNDYVAPVGVKFCDFGFIPIISILGELITMRQIKKKRATIDPIBK 480  
Db 778 VSPDTLREGCKEYDVAPOVGHRCDFGFIPIISILGELITMRQIKKKRATIDPIBK 837  
QY 481 MLDYRQAVKLHA 493  
Db 838 LLDYRQAVKLHA 897  
QY 494 493  
Db 898 VSGLEVPSENRRTNKAELKRVKALIRHDYSGKVTYIRLKSGRKITSHSLFSVRNGEL 957  
QY 494 493  
Db 958 VEVTDGLKPGDLVAVPRELEPERNHVNLVLLGTPEETLDMVTIPVKGKKNFFK 1017  
QY 494 493  
Db 1018 GMLRTLWIFGEKRPRTARRYLRLHEDLGYVRLKKGIVYLDWLSKNRYLYALVEN 1077  
QY 494 493  
Db 1078 VRYNGNKBYLVFENSIRDAVGIMPLKELKWKIGTLNGFRMRKLIIEVDES LAKLLGYV 1137  
QY 494 493  
Db 1138 SEGARYKQNPKNKNGWSYVKLYNEDPEVLDDEMLASRFFGKVRGRNRYVEIPKIGYLL 1197  
QY 494 493  
Db 1198 FENMCGVLAENKRIPEFVFTSPKGVLAFLGFIYGDGDVHNPKRILSTKSELLANQV 1257

QY 494 493  
Db 1258 LLLNSGVSAVLGHDGSGVYRVYINBELPFVKLDKKKNAYSHVIPKEVLSEVFGKVFQK 1317  
QY 494 493  
Db 1318 NVSPQIFRKWVEDGRDLDPKQAQRSLWLBEGDVVLDRVESVDYDGYVYDLSVEDNEF 1377  
QY 494 493  
Db 1378 LVGGLVYAHNSYGYGYARAEWYCKEASVTAWGREYITMTIKIEBEKYGFKVIYSD 1437  
QY 544 TDGFYATIPCEKPEKTIKKAKELKYNKSLPGLLELEYEGFYLRGFFVAKKXAVYIDEE 603  
Db 1438 TDGFFATIPGADAETVKKKAMEFLKYNKSLPGLLELEYEGFYLRGFFVAKKXAVYIDEE 1497  
QY 604 GRITTRGLEVRDRWSEIAKETQAKVLEAILKEDSVKEAVEIKDVVEEIAKYQVPLEKL 663  
Db 1498 GKITTRGLEVRDRWSEIAKETQAKVLEAILKEDSVKEAVEIKDVVEEIAKYQVPLEKL 1557  
QY 664 VIHQITKDLSEYKATGPHVAIAKLAAGIKVYRPGTIIISYIVLGRGKISDRVILLSEY 723  
Db 1558 VIHQITRDLKDYKATGPHVAIAKLAAGIKVYRPGTIIISYIVLGRGKISDRVILLSEY 1617  
QY 724 DPXKHYPDPDYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWLK 773  
Db 1618 DPTKHKYDAEYIENQVLPVLRILEAFGYRKEDLKYQSSKQVGLDAWLK 1667  
RESULT 11  
DPOL\_THEG8 STANDARD; PRT; 1699 AA.  
AC QHH84;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA polymerase (EC 2.7.7.7) [Contains: Endonuclease PI-Tspge8I  
DE (EC 3.1.1.-) (Tsp-GE8 pol-1 intein); Endonuclease PI-Tspge8II  
DE (EC 3.1.1.-) (Tsp-GE8 pol-2 intein)].  
GN POL OR POL-1.  
OS Thermococcus sp. (strain GE8).  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Thermococcus.  
OX NCBI\_TaxID=105583;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Querrelou J.J.E., Cambon M.A., Lesongeur F., Barbier G.;  
RT "Thermococcales taxonomy and phylogeny based on the comparative use of  
RT 16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase  
RT genes";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
CC EXHIBITS 3', TO 5' EXONUCLEASE ACTIVITY (BY SIMILARITY).  
CC -!- FUNCTION: PI-TSPGE8I AND PI-TSPGE8II ARE ENDONUCLEASES  
CC (POTENTIAL).  
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA}[N].  
CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEN).  
CC FOLLOWED BY PEPTIDE LIGATION.  
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
CC -!- SIMILARITY: IN THE INTEN SECTION; BELONGS TO THE HOMING  
CC ENDONUCLEASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC





QY 697 REGTISIVLGGSKISDRVLLSEYDPKHKYDPDYIYENOVLPVLRILEAFGYRKE 756  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1618 REGTISIVLGGSGRIGRAIPFFDFDFAKHDAEYIYENOVLPVLRILEAFGYRKE 1677  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 QY 757 DLKYSQSSKOVGLDAMLK 773  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1678 DLRYQKTKOVGLGAWLK 1694  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

## RESULT 12

DPOL THEHY  
 ID QPOL THEHY STANDARD; PRT; 1668 AA.  
 AC Q9HR05;  
 DT 16-OCT-2001 (rel. 40, Created)  
 DT 16-OCT-2001 (rel. 40, Last sequence update)  
 DT 28-FEB-2003 (rel. 41, Last annotation update)  
 DE DNA polymerase (EC 2.7.7.7) [Contains: Endonuclease PI-ThyII  
 (EC 3.1.-.-) (Thy pol-1 intein); Endonuclease PI-ThyI (EC 3.1.-.-)  
 (Thy pol-2 intein)] (fragment).  
 GN POL.  
 OS Thermococcus hydrothermalis.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 CC Thermococcus.  
 CC NCBI\_TaxID=46539;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Querellou J.E., Cambon M.A., Lesongeur F.O., Barbier G.;  
 RT "Thermococcales taxonomy and phylogeny based on the comparative use of  
 RT 16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase  
 RT genes."; Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN CHARACTERIZATION OF PI-THYI.  
 RP MEDLINE=20512590; PubMed=11058140;  
 RA Saves I., Bleau H., Dietrich J., Masson J.-M.;  
 RT "The Thy pol-2 intein of Thermococcus hydrothermalis is an  
 RT isochizomer of PI-ThyI and PI-TfuII endonucleases."; Nucleic  
 RL Nucleic Acids Res. 28:4391-4396 (2000).  
 CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
 CC EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY (BY SIMILARITY).  
 CC -!- FUNCTION: PI-THYI AND PI-THYII ARE ENDONUCLEASES. PI-THYI CLEAVES  
 CC THE INTEINLESS SEQUENCE OF THE THY DNA POL. GENE. IT REQUIRES A 21-  
 CC BP MINIMAL RECOGNITION SEQUENCE.  
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + [DNA](N).  
 CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES  
 CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)  
 CC FOLLOWED BY PEPTIDE LIGATION (potential).  
 CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
 CC -!- SIMILARITY: IN THE INTEIN SECTION, BELONGS TO THE HOMING  
 CC ENDONUCLEASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ245619; CAC18555.1; ..  
 CC HSSP; P56689; ITGO.  
 CC REBASE; 4832; PI-ThyI.  
 CC InterPro; IPR006172; DNA pol B.  
 CC InterPro; IPR006134; DNA pol B.  
 CC InterPro; IPR006133; DNA pol B.  
 CC InterPro; IPR003866; Hedgehog HintC.  
 CC InterPro; IPR003587; Hedgehog HintC.  
 CC InterPro; IPR006141; Intein.  
 CC InterPro; IPR006142; Intein.  
 CC InterPro; IPR004042; Intein.  
 CC InterPro; IPR004578; Pol2.  
 CC Pfam; PF00136; DNA\_pol\_B; 3.

DR Pfam; PF03104; DNA\_pol\_B\_exo; 1.  
 DR PRINTS; PRO0379; INTEIN.  
 DR SMART; SM00305; HintC; 2.  
 DR SMART; SM00306; HintN; 2.  
 DR SMART; SM00486; POLBc; 1.  
 DR TIGRfams; TIGR01443; intein\_Cterm; 2.  
 DR TIGRfams; TIGR01445; intein\_Nterm; 2.  
 DR TIGRfams; TIGR00592; pol2; 2.  
 DR PROSITE; PS00116; DNA POLYMERASE\_B; FALSE\_NEG.  
 DR PROSITE; PS00818; INTEIN\_C\_TER; 2.  
 DR PROSITE; PS00819; INTEIN\_ENDONUCLEASE; 2.  
 DR PROSITE; PS00817; INTEIN\_N\_TER; 2.  
 KW Transferase; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; Hydrolase; Nuclease; Exonuclease; Multifunctional enzyme;  
 KW Protein splicing; Autocatalytic cleavage; Endonuclease; Intrin homing.  
 FT CHAIN 1  
 FT CHAIN 1 458 DNA POLYMERASE, 1ST PART.  
 FT CHAIN 459 995 ENDONUCLEASE PI-THYI (POTENTIAL).  
 FT CHAIN 996 1044 DNA POLYMERASE, 2ND PART.  
 FT CHAIN 1045 1433 ENDONUCLEASE PI-THYII (POTENTIAL).  
 FT CHAIN 1434 1668 DNA POLYMERASE, 3RD PART.  
 SQ SEQUENCE 1668 AA; 193319 MW; 5EEB005FEFDA71C8 CRC64;  
 Query Match 63.0%; Score 2535.5; DB 1; Length 1668;  
 Best Local Similarity 33.6%; Pred. No. 3.3e-130;  
 Matches 560; Conservative 85; Mismatches 92; Indels 929; Gaps 4;  
 QY 34 FQYIVALLKDDSAIDKRAIGERHGKIVRVVDVAVKVKKTLGRDVEYWKLIFFHPQDV 93  
 Db 1 FEYIVALLKDDSAIDKRAIGERHGKIVRVVDVAVKVKKTLGRDVEYWKLIFFHPQDV 60  
 QY 94 PALRGKIREHPAVIDIYEDIPFAKYLIDKGLIMPGEDEELKMAFDITFTYHEGDEFG 153  
 Db 61 PAIRDEIRRHSAVVDIYEDIPFAKYLIDKGLIMPGEDEELKMAFDITFTYHEGDEFG 120  
 QY 154 KGEIIMISYADEEAEARVITWKIDIPYDVVNEREMIKFVQIVREKDPDLITNGDN 213  
 Db 121 TGPILMISYADEEAEARVITWKIDIPYDVVNEREMIKFVQIVREKDPDLITNGDN 180  
 QY 214 FDLFYLIKRAEKLGVTLGRDKEHPKIHMGSGSFAYEIKRIFHDFLPVVRRTINLP 273  
 Db 181 FDFAYLKKECKIGIKFTLRDGS--EPKIQMGDRFAVEVKGRIHFDLPVVRRTINLP 238  
 QY 274 TYTLEAVYEAIVGKTSKLGABEIAAIWETESMKLAQYSMEDARATVELGKEPPEMEA 333  
 Db 239 TYTLEAVYEAIVGKTSKLGABEIAAIWETESMKLAQYSMEDARATVELGKEPPEMEA 298  
 QY 334 ELAKLIGQSVDSRSSTGNLVEWYLLRVAYERNELAPNKPDEEYRRRLRTTYLGGYVK 393  
 Db 299 QLSRLIGQSVDSRSSTGNLVEWYLLRVAYERNELAPNKPDEEYRRRLRTTYLGGYVK 357  
 QY 394 EPERGLWENITYLDFCLPSYIIVTHNVSPDTLREGCKQYDVAPIVGKFKDPFGFIP 453  
 Db 358 EPERGLWENITYLDFCLPSYIIVTHNVSPDTLREGCKQYDVAPIVGKFKDPFGFIP 417  
 QY 454 SILGELITVRQETIKKKKATIDPIEKMKLDYQRAVKLHA----- 493  
 Db 418 SLGALLDERQKIKKKKASIDPLEKLDYQRAVKLHA----- 477  
 QY 494 ----- 493  
 Db 478 RICEFVVKLMDTSDLVKNGDTEVLEVRGIRALSDFRASKKARVMFKAIRHRYSGDV 537  
 QY 494 ----- 493  
 Db 538 YEIVLGSRRITVTEGHSIFAYGDGELREVGTGKAGDALLAVPRVNLPEKKERLNLYE 597  
 QY 494 ----- 493  
 Db 598 LLRLPEETGDIILTIPVKGKKNPFKMLRTLRLWISGEEKRPTARRYLEHLEGLGYVR 657  
 QY 494 ----- 493



```

Db 658 LKKIGYVTDREGLRYKLYERLVEAVRYNGKREYLVEFNAVRDVIALMPBEELRDWL 717
QY 494 ----- 493
Db 718 VGTFRGFRMPFVEIBEDFAKLLGYVVBGNARKWENKNGSYTVKLYNQRVLDDME 777
QY 494 ----- 493
Db 778 SLAERFFRVKRGKNIETPRKWAYIIFENLCOTLAENKRVPAIFTSPEVRFVAFIEGY 837
QY 494 ----- 493
Db 838 FIGDGVHPSKRVRLSTKSELLVNLGLVLLNSLGVSAIKIRHDSGVYRVYVNEELPPTDY 897
QY 494 ----- 493
Db 898 RKKQNAVYSHVPIKEIILEETFGKVFQFSYERFRELKSEKLDGKAKRIEVLNGDVV 957
QY 494 ----- 493
Db 958 LDKVLEVKKRPYEGYVVDLSVEEDENFLAGFLLYAHNSYGYGYARARWYCKECAESV 1017
QY 517 TANGRHYIEMTIKIEBKFGKVLAD----- 543
Db 1018 TANGRDYIETTHIEIERFGKVLADSVTGETETETETETETETETETETETETETET 1077
QY 544 ----- 543
Db 1078 EKECVLEGVEALTDNRGLVWKSVPYVMRHTNKRIYRWNTSNWYLDVTDHSLIGY 1137
QY 544 ----- 543
Db 1138 MNTSKVPGKPKERLVEVKPGELGESVKSLITDNRAIAHGIRVNPVIAVKLWELLGLLVG 1197
QY 544 ----- 543
Db 1198 DGNWGGOSNAKYVNLGLDKDEIEBEKILPKLNTGIISNYDKSKGDVLSILSKWLA 1257
QY 544 ----- 543
Db 1258 RPYRYFKDESGSKRIPEFMFNLPREYIEAFLRGLFSADGTVSLRGVPEVRLTSVNP 1317
QY 544 ----- 543
Db 1318 SSSVRKLLWLVGVNSMFMETPNPNRYLGKESGTHSVHVRKDKHFAERIGFLDLDRKATK 1377
QY 544 ----- 543
Db 1378 LSENLGHTSKKRAYKYDFDLVYPKVEEIAVDGVVYDIEVEGTHRFANGILVHNTDGF 1437
QY 548 YATIPGEKPEPTIKKAKAEFLKINSKLPGLLELEYEGHYLGGFVAKRYAVIDEGRIT 607
Db 1438 FATIPGADAETVKKAKAEFLKYNKAKLPGLLELEYEGHYVGRFFVTKKXAVIDEGRIT 1497
QY 608 TGLGVWRDSEIAKETQAKVLEAILKEDSVKRAVEIKVDWBEIEAKYQVPLEKVIHE 667
Db 1498 TGLGVWRDSEIAKETQAKVLEAILKEDSVKRAVEIKVDWBEIEAKYQVPLEKVIHE 1557
QY 668 QITKDLSEYKAIGPHVAIAKLAAGIKVRGTTISYIVLKGSKISDRVILLSEYDPKK 727
Db 1558 QITRELKDYKATGPHVAIAKLAAGIKVRGTTISYIVLKGSKISDRVILLSEYDPKK 1617
QY 728 HKYDPTYIENQVLPAVLRIIEAFGYRKEDLYQSSKQVGLDAMLK 773
Db 1618 HRYDAEYIENQVLPAVLRIIEAFGYRKEDLYQSSKQVGLDAMLK 1663

RESULT 13
DPOL.METVO
ID DPOL.METVO STANDARD; PRT; 824 AA.
AC P52025;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL.
OS Methanococcus voltae.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2188;
RN [1].
RP SEQUENCE FROM N.A.
RX MEDLINE=95014087; PubMed=7929013;
RA Konisky J., Paule S.M., Carinato M.E., Kansy J.W.;
RT "The DNA polymerase gene from the methanogenic archaeon Methanococcus
RT voltae.";
RL J. Bacteriol. 176:6402-6403 (1994).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L33366; AAA72443.1; -.
DR HSP; P56689; 1TGO.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR InterPro; IPR004578; Pol2.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNAPOLE.
DR SMART; SM00486; POLB; 1.
DR TIGRfam; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding.
SQ SEQUENCE 824 AA; 96754 MW; 094A630D8C1514FC CRC64;

Query Match 34.6%; Score 1392; DB 1; Length 824;
Best Local Similarity 36.1%; Pred. No. 1.9e-68;
Matches 308; Conservative 179; Mismatches 251; Indels 116; Gaps 21;

QY 4 DTDYTKGKPIRIFFKENG-----EFKIELDPHQPIYIALLKDDSAI-----D 49
Db 2 DLDYNSKD--LCIDMYKNCGLKKEPEINLQKECFKPYVDTSEPKIYDYLGLNOEI 59
QY 50 EIKATKGERHG---KIVRVVDVAVKVKKFL-----GRDV-----EYWKLI 86
Db 60 DLKLEPEPENTSLKVQDLITNIEIKIYVSDVILNGKDISEVDFNKKERICKVY 119
QY 87 FEHPQDVPALRKIEHPAVIDIYDYDPPFAKRYLIDKGLIPM-----EGDEELK 136
Db 120 VKYPNHVKIIRYFKEFG---KSYEFDPFLRRYIMIDQDIVPSAKYSDNKKIDNSIP 176
QY 137 LMAFDLETYHSGDFEGKGEIIMISYADBEARVITWKNIDLP-----YDVVVSNEREMIK 192
Db 177 CLAFDNELCKEPEAKKADPIIWNLFSDYQKVIYTKFENSEYNGCVDYVYKDEKLIQ 236
QY 193 RFVQIVREKDPVLITNGDNFDLPYLKRAEKLGVTLGLGRDKHEP---KIRMGDS 249
Db 237 KTEIL--KQYDVITYNGDNFDPYLKRAIYIELDFDNASNSQQPQIIKISGGIN 294
QY 250 FAVEIKGRTHFDLPVVRRTINLPTVLEAVVGLTKSKLGAEEIAAIVETESMECK 309
Db 295 RKSKIPIGIHIDLYPIARKLLNLTVKLENVQELFKINKEAVDYGDPKMWETEDT--T 352
QY 310 LAQYSMEDARATVELGKEFFPMEAEIAKLIGOSVMDVSRSSTGNIVYVLLVAVERNEL 369
Db 353 LLRAYVEDALYTYMGNYVFLPLEIMPSRVNQNPPLYDTSRMNSQWVEFLLKRSPEQNM 412

```



```

QY 690 AAKGKVRPCTIISYIVLRSGKISDR---VILLSEYDPKK-----HKYDPDY 735
D 686 REMGIIYVSSKIGYIVKSGNIGRAVPIDLIEDFGENLRKTKSGIEIKLKDXY 745
QY 736 IENQVLPVLRILIEARGYKEDLYKQSSKQVGLDAW 771
D 746 IDNQIIPVLRILIEARGYKEDLYKQSSKQVGLDAW 779

RESULT 15
DPOL METJA STANDARD; PRT; 1634 AA.
AC Q58295;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) [contains: Mja pol-1 intein; Mja pol-2
DE intein].
GN POL OR MJ0885.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geohagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Corton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -!- PGM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).

```

```

DR TIGRfams; TIGR01445; intein_Nterm; 2.
DR TIGRfams; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA POLYMERASE_B; 1.
DR PROSITE; PS00818; INTEIN_C_TER; 2.
DR PROSITE; PS00819; INTEIN_ENDONUCLEASE; 2.
DR PROSITE; PS00817; INTEIN_N_TER; 2.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Autocatalytic cleavage; Protein splicing;
KW Complete proteome.
FT CHAIN 1 425 POL, 1ST PART (POTENTIAL).
FT CHAIN 426 794 MJA POL-1 INTEIN (POTENTIAL).
FT CHAIN 795 882 POL, 2ND PART (POTENTIAL).
FT CHAIN 883 1358 MJA POL-2 INTEIN (POTENTIAL).
FT CHAIN 1359 1634 POL, 3RD PART (POTENTIAL).
SQ SEQUENCE 1634 AA; 191708-MW; 841FAFAB1F97DDD CRC64;

Query Match 38.1%; Score 1292; DB 1; Length 1634;
Best Local Similarity 22.7%; Pred. No. 1.3e-63;
Matches 374; Conservative 159; Mismatches 217; Indels 894; Gaps 23;

QY 2 IFDIDYITKGGKPIIRIKKENGGEFKIELDPHFOPYIYALK---DSDAIDEIK--AIK 55
D 13 LIDNTYKTIEDKAVIYLINS---ILKRDFFKPYFVELHKEKVENEDIEKIEKELLK 68
QY 56 GERHGKIVRVVDVAVKVKKK-FLGRDVEVWKLIFEPHQDVPALRGKIREHPAVIDIYEVDI 114
D 69 ND---LLKPVENIEVWVKIILRKEKEVIKIIATHPKQPKLR-KIKECEIVKEIYEHDI 123
QY 115 PFAKYLDKGLIPM---EGDE-----ELKLMADIDETVHEGDEFGKE-IIMIS 161
D 124 PFAKYRLIDNIIIPMTYDFENKPKVPSEIPLKLSVAFDMEVYTRDTPNPERDILMAS 183
QY 162 YADDEARVITWKNIDLPVDVNSNEREMIKRFYQIVREKDPDYLITYNGDNFDPYLIK 221
D 184 FWDENGKGVITVYEFNHNIEVWVKNEKELIKIITETKEY--DVIYTYNGDNFPPYLKA 241
QY 222 RAEKLGVTLLIGRDKHEPEPKIHMGDSFAVEIKGRITHFDLPVVRRTINLPTYLEAVY 281
D 242 RAKYIGIDINLGRDGE--ELKIKGGMEYRSYIPGRVHIDIDLPISRLLKUTKVTLEDVV 299
QY 282 EAVLGKTKSKLGABEIAAIWETESMKLAQYSEMEDARATYELGKEFFPMAELAKLIGQ 341
D 300 YNLFGIEKLIKPHTKIVDYWANN--KTLIEYSIQDAKYTYKIGYFPFELVMSRIWQ 357
QY 342 SWDVSRSTGNLWVWYLLRVAYERNELAPKNPBESEYRRLRTTYLGGYKKEPERGLWE 401
D 359 TPFEITRMSGGQWBYLLMKAFKAFENMIVPNKPDSEYRRLVLTTEGGYKKEPEKGMFE 417
QY 402 NITYLDPRC----- 410
D 418 DIISMDFRCHPKGTWVWVGKGINVEDYKGNVVLGIDGKQKVKWKVYEGEELINVN 477
QY 411 ----- 410
D 478 GLKCTPNHKIPURYKIKHKKNKNDYLRDIYAKSLLPKSGKLLCKDPETIGNYEK 537
QY 411 ----- 410
D 538 YINDMDEDEILKSELIGILLAEGLHLLRRDIEYFDSRSGKKAISHQRYVEITWNEDEKDFI 597
QY 411 ----- 410
D 598 EKIKYIFKKLFNYLYVRKKGTKAITLGCACKDIYLIKIEILKNKYEKLPNAILRGPF 657
QY 411 ----- 410
D 658 GGGVNTVRVAVVVGQNTNNDYKIKFIASLLDLRLGIKYSFYTSYSEERGGKLLKRYVIF 717
QY 411 ----- 410
D 718 SKGDLIKFSILISFTSRKKNLNLNIEIROKLYKIGDYFDLDVDCVSLSEYKGEVYDL 777
QY 411 -----LYPSIIIVHNVSPTLREGCKNYDVA-PIVGYKFKCDFG 451

```